Untangling the STRUCTURE *and* DYNAMICS *of ecological networks*

Bernat Bramon Mora

June 1, 2019

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
at the University of Canterbury

School of Biological Sciences
University of Canterbury
New Zealand

Bernat Bramon Mora: *Untangling the Structure and Dynamics of Ecological Networks;* Developing computational tools for the study of ecological communities; © June 1, 2019

La vida sovint ens porta a repetir postures interessants.
Una d'elles—la meva preferida—és la del primer bany de l'any;
una persona mirant mar enllà amb l'aigua als genolls i les mans a la cintura,
sense cap altra alternativa que la de llençar-se al mar tard o d'hora.

— L'home del principi

ABSTRACT

Networks provide a simple way to synthesize the heterogeneous nature of ecological systems. They depict ecological communities as graphs of species and their interactions, and they have been extensively used in ecology to study topics such as community stability, biodiversity, and global change. However, untangling the complexity encoded within these objects is far from straightforward. Ecological networks can encompass many levels of information, and their analysis is inherently challenging. One way ecologists have tried to do this is by comparing ecological networks to random models and across ecosystems or environmental gradients. With the right set of tools, one can shed light on different aspects regarding the structure and dynamics of ecological communities.

This thesis presents a variety of statistical tools designed for the study of complex networks, which my co-authors and I put to work on a diverse set of ecological datasets to answer questions relevant to some longstanding topics of community ecology. The first chapter provides a method for generating "correlation-informed" null models, a general and flexible approach to study the mechanisms explaining such structures. Comparing these models to empirical data, we assess whether the information encoded within different correlation structures is predictive for explaining network patterns commonly used to study ecological communities. In the second chapter, we instead compare networks across different ecosystems using a novel network-comparison techniques—network alignment—that allows us to map ecological communities on top of each other and study their structural similarity. The idea behind network alignment is to pair up species from different communities based on the "roles" these play in their respective systems, which can be based on any species' property. Following this, the third chapter introduces a new way of characterizing these species' roles that accounts for the strengths of the interactions between species. Finally, we present a conceptual framework in the fourth chapter that combines the tools presented throughout this thesis, providing us with a new way to analyze network time series. Focusing on the seasonal dynamics of ecological networks, we use this framework to break down the dynamics of the interactions between plant and pollinator species.

ACKNOWLEDGMENTS

After four years, there is undoubtedly many people that I would have to thank for their help and support, both technical and emotional. That said, I have seen too many theses with a cheesy acknowledgments section, and I am going to keep this one brief. I am aware that this is often the most read part of a thesis. If you are here looking for your name, I would like to apologize in advance because you are likely going to be disappointed.

I would like to thank the following people: Daniel, who has been an amazing supervisor and mentor; my family and Eleonore for their unconditional love and support; Fer, Marilia and Matt for a lifetime friendship; the current and former members of the Stouffer and Tylianakis labs; and any friend that have directly or indirectly helped me get where I am.

* It is not often that one feels at home in the antipodes.

PREFACE

This thesis has been written as cohesive set of stand-alone scientific articles, all of which share a common focus: developing statistical tools for the study of ecological networks. At the time of the submission of this thesis, each of these articles were in different stages of the publication process. The first chapter, "Unmasking structural patterns in incidence matrices: an application to ecological data", was published in the *Journal of the Royal Society Interface* February 2019: volume 16, number 151, pages 20180747. The second, "Identifying a common backbone of interactions underlying food webs from different ecosystems", was published in *Nature Communications* July 2018: volume 9, number 1, pages 2603. The third, "pymfinder: a tool for the motif analysis of binary and quantitative complex networks", was pending resubmission to *PLOS Computational Biology*. The fourth, "Untangling the seasonal dynamics of plant-pollinator communities", was in preparation for submission to *Ecology Letters*. The first appendix, "Seeing the forest for the trees: putting multilayer networks to work for community ecology" was published in *Functional Ecology* November 2018: volume 33, number 2, pages 206–217. Finally, the "general introduction" for this thesis connects the different chapters, providing a broad context for them. The "general discussion", on the other hand, expands beyond the ideas presented in the different chapters, discussing some of the potential steps moving forward.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Part I

GENERAL INTRODUCTION

ECOLOGICAL NETWORKS

> *"Now diversity is of theoretical interest because it can be related to stability, maturity, productivity, evolutionary time, predation pressure, and spatial heterogeneity... [but] diversities are mere numbers and should be distinguished from the theories which they support"* (Hill, 1973).

Hill was speaking to the rise of diversity measures in Ecology in the early 70s. At the time, many ecologists were uneasy with such measures, confused by their ambiguity and finding richness, the direct count of species, a much more intuitive concept (Hurlbert, 1971). Diversity measures, such as Shannon's entropy or evenness, ultimately proved useful because they contain more information than absolute measures of species richness; they not only describe which species are present, but also in what relative amounts. Hill decides to conclude, however, with an insightful message. These diversity measures are just useful tools that we use to study ecological communities; they should just be treated as a means to an end. The same rational applies to a modern parallel: ecological networks.

Any individual organism exists not despite but because of its interspecific interactions with its ecological community. It is not a surprise, therefore, that network science has found the perfect culture medium in ecology, and that the concept of a network has been able to blossom into the ecological mainstream over the past 20 years. In mathematics, a network—or a graph—is simply a representation of a system using a collection of nodes connected by links (Newman, 2003). In ecology, however, these nodes represent living organisms, and links capture the intricate relationships between them (Cohen et al., 2012).

Networks provide a simple way to synthesizing the natural complexity underlaying ecological systems. Indeed, these objects can take a wide variety of forms. For example, the link between two nodes in a network can be designed to represent the feeding interaction between a starfish and a mussel in a marine community (Paine, 1966), the mutualistic interaction between a hummingbird and a flower in a tropical forest (Feinsinger, 1978), or even the parasitic relationship between a tick and a mammal species (Wells et al., 2013). Likewise, networks can describe ecological communities at very different scales, from small food webs formed in water-filled tree-holes (Kitching, 1971) to large ecological networks comprising different interaction types (Melián et al., 2009) or spanning over multiple years (Burkle et al., 2013). In

this context, networks can be crucially put to work to understand dynamical stability of species (May, 1972) as well as speak to heart of central topics in ecology such as biodiversity (Dunne et al., 2002b; Bascompte et al., 2006) and global change (Tylianakis et al., 2008). Much like the old debates regarding diversity measures, however, one must keep in mind that ecological networks are simply useful tools. Ecological studies should therefore treat them as such.

Most ecological studies that take a network approach tend to focus on one of two aspects: structure and dynamics. The structure of ecological networks, on the one hand, has been shown to be very non-random (Briand and Cohen, 1984; Dunne et al., 2002a), and understanding the eco-evolutionary mechanisms non-randomly shaping communities across environments has been a central challenge in ecology (Tylianakis and Morris, 2017). Unfortunately, the structural studies of ecological networks have at times been overly descriptive, with some arguing that the field would benefit if networks were instead used as investigative tools in hypothesis-driven research (Blüthgen, 2010; Poisot et al., 2016). On the other side, studies on the dynamics of ecological communities address questions about community stability and species' persistence to perturbations (Dunne et al., 2002b, 2004; Stouffer and Bascompte, 2011). However, the dynamics of complex networks is often studied through the lenses of complex mathematical models (Bastolla et al., 2009). Although such models have been very successful at linking certain aspects of the structure of ecological systems to their dynamics (Bascompte et al., 2003; Thébault and Fontaine, 2010), they generally rely on broad mathematical simplifications that might critically compromise the ecological realism of the results (Yodzis, 1998; Holland et al., 2002).

Throughout this thesis, I develop multiple computational tools for the study of structure and dynamics of ecological communities. In each case, I put such tools to work on a diverse set of ecological datasets, answering questions that are relevant to some longstanding topics of community ecology. I start here with the study of common network metrics. Network metrics have often been used by ecologists as a way to characterize the structure of ecological communities (Dunne et al., 2002a). Examples include relatively short food-chains (Lawton, 1989) and a roughly constant fraction of top, intermediate, and basal species(Hall and Raffaelli, 1993) in antagonistic communities, and a nested pattern found in the way plant-pollinator interactions are distributed within mutualistic networks (Bascompte et al., 2003). That said, the single measure of a particular network metric is not necessarily informative—a network metric alone can tell you very little about the system unless this is compared to "something". Following

this, I begin this thesis with a fresh look at an old concept in community ecology: null models.

Null models are crucial statistical tools that have been extensively used in modern ecology and biogeography to test different hypotheses regarding phenomena such as community assembly (Gotelli, 2000; Gotelli and Entsminger, 2001; Manly, 2006; Gotelli and Ulrich, 2010; Miller et al., 2016). These models are often based around randomizations of ecological data that, when compared to empirical data, can provide insights into how the data is structured and the potential mechanisms explaining such structure (Gotelli, 2001). **In the first chapter of this thesis**, my co-authors and I worked on ways to inform null models for the study of the mechanisms explaining some common network metrics used in network science. With this approach, we studied the phylogenetic component underlying trophic network structure and the factors behind one of the most used patterns in island biogeography studies.

Our approach offers general and flexible way to study the mechanisms explaining the structure of biological communities. However, analyzing the structure of networks using network metrics has fundamental limitations. These descriptors are summary statistics at the network level and mostly overlook the actual way ecological interactions are distributed within a network. As a result, these metrics can often mask more subtle—but no less ecologically important—variation across communities. **In the second chapter of this thesis**, my co-authors and I offer an alternative approach that overcomes this shortcoming by directly aligning pairs of networks to each other. We apply this alignment approach to investigate the topological similarity of a large collection of food webs—networks of prey-predator relationships between species—from multiple environments and biomes. Perhaps most importantly, we study the differences in the way species from different type of ecological communities interact with each other, testing previous suggestions of universality across food webs.

The idea behind network alignment is to map networks on top of each other by pairing their components based on the *roles* these play in their respective systems. Notice that the term *role* is purposely ambiguous here since it can describe any property of the nodes' nature. This ambiguity makes the alignment approach very versatile because it allows us to compare networks based on any species' properties. **In the third chapter of this thesis**, my co-authors and I develop a new way to characterize species' structural roles. Our definition of a species' role in a network is an extension of the idea of motif roles defined by Stouffer et al. (2012), but it also incorporates information about the interaction strengths between species. This is important because the interaction strengths are a critical aspect of certain ecologi-

cal communities, and they have been shown to be key to understand certain aspects of community dynamics (McCann et al., 1998; Kokkoris et al., 1999; Bascompte et al., 2006).

Combining the new definition of a "weighted motif role" and the alignment method developed in the second chapter, we then turn our focus to the study of community dynamics. Ecological communities are inherently dynamic, and synthesizing such dynamics into a general framework has become one of the key challenges in ecology (Pellissier et al., 2018). Perhaps one of the main difficulties to doing so is finding the "right" scale for this synthesis. **In the fourth chapter of this thesis**, my co-authors and I show how to find a middle ground between 'full-network' and 'single-species' approaches to untangle the seasonal dynamics of plant-pollinator networks. In particular, we define the concept of a species' position within a network, and use this concept to understand how species change their positions over time in network time-series. While the concept of a species' role depends on the way a species interacts relative to all the others, the position of a species depends instead on the roles of all the other species. Following this, we identify groups of distinct positions across networks and condense the complex dynamics of individual species over time into something much simpler: the movement of species across positions. As a result, we are able to distill some general rules regarding species' interaction turnover, phenology, and assembly processes in empirical plant-pollinator communities.

Finally, in an appendix following the main body of the thesis, I present additional work carried out during my PhD candidature at the University of Canterbury. This work is the result of a symposium that I co-organized with Matthew C. Hutchinson and Daniel B. Stouffer at the 2017 Ecological Society of America Annual Meeting in which we discussed the concept of multilayer networks, central to some of the ideas presented in this thesis. In the appendix, my co-authors and I review some of the challenges and possibilities of using multilayer networks to study the structure of ecological communities. We reflect on several aspects regarding the empirical implementation of ecological multilayer networks, and we outline the different research areas that could most immediately benefit from such an approach.

BIBLIOGRAPHY

Jordi Bascompte, Pedro Jordano, Carlos J Melián, and Jens M Olesen. The nested assembly of plant–animal mutualistic networks. *Proceedings of the National Academy of Sciences*, 100(16):9383–9387, 2003.

Jordi Bascompte, Pedro Jordano, and Jens M Olesen. Asymmetric coevolutionary networks facilitate biodiversity maintenance. *Science*, 312(5772):431–433, 2006.

Ugo Bastolla, Miguel A Fortuna, Alberto Pascual-García, Antonio Ferrera, Bartolo Luque, and Jordi Bascompte. The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature*, 458(7241):1018, 2009.

Nico Blüthgen. Why network analysis is often disconnected from community ecology: a critique and an ecologist's guide. *Basic and Applied Ecology*, 11(3):185–195, 2010.

Frédéric Briand and Joel E Cohen. Community food webs have scale-invariant structure. *Nature*, 307(5948):264, 1984.

Laura A Burkle, John C Marlin, and Tiffany M Knight. Plant-pollinator interactions over 120 years: loss of species, co-occurrence, and function. *Science*, 339(6127):1611–1615, 2013.

J. E. Cohen, F. Briand, and C. M. Newman. *Community food webs: data and theory*, volume 20. Springer Science & Business Media, 2012.

Jennifer A Dunne, Richard J Williams, and Neo D Martinez. Food-web structure and network theory: the role of connectance and size. *Proceedings of the National Academy of Sciences*, 99(20):12917–12922, 2002a.

Jennifer A Dunne, Richard J Williams, and Neo D Martinez. Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecology Letters*, 5(4):558–567, 2002b.

Jennifer A Dunne, Richard J Williams, and Neo D Martinez. Network structure and robustness of marine food webs. *Marine Ecology Progress Series*, 273:291–302, 2004.

Peter Feinsinger. Ecological interactions between plants and hummingbirds in a successional tropical community. *Ecological Mono-*

graphs, 48(3):269–287, 1978.

Nicholas J Gotelli. Null model analysis of species co-occurrence patterns. *Ecology*, 81(9):2606–2621, 2000.

Nicholas J Gotelli. Research frontiers in null model analysis. *Global Ecology and Biogeography*, 10(4):337–343, 2001.

Nicholas J Gotelli and Gary L Entsminger. Swap and fill algorithms in null model analysis: rethinking the knight's tour. *Oecologia*, 129 (2):281–291, 2001.

Nicholas J Gotelli and Werner Ulrich. The empirical Bayes approach as a tool to identify non-random species associations. *Oecologia*, 162(2):463–477, 2010.

Stephen J Hall and Dave G Raffaelli. *Food webs: Theory and Reality*, volume 24. Academic Press, 1993.

Mark O Hill. Diversity and evenness: a unifying notation and its consequences. *Ecology*, 54(2):427–432, 1973.

Julian N Holland, Donald L DeAngelis, and Judith L Bronstein. Population dynamics and mutualism: functional responses of benefits and costs. *The American Naturalist*, 159(3):231–244, 2002.

Stuart H Hurlbert. The nonconcept of species diversity: a critique and alternative parameters. *Ecology*, 52(4):577–586, 1971.

RL Kitching. An ecological study of water-filled tree-holes and their position in the woodland ecosystem. *Journal of Animal Ecology*, pages 281–302, 1971.

GD Kokkoris, AY Troumbis, and JH Lawton. Patterns of species interaction strength in assembled theoretical competition communities. *Ecology Letters*, 2(2):70–74, 1999.

John H Lawton. *Ecological Concepts*, chapter Food webs, pages 43–78. Blackwell Scientific, 1989.

Bryan FJ Manly. *Randomization, bootstrap and Monte Carlo methods in biology*, volume 70. CRC Press, 2006.

Robert M May. Will a large complex system be stable? *Nature*, 238 (5364):413, 1972.

Kevin McCann, Alan Hastings, and Gary R Huxel. Weak trophic interactions and the balance of nature. *Nature*, 395(6704):794, 1998.

Carlos J Melián, Jordi Bascompte, Pedro Jordano, and Vlastimil Krivan. Diversity in a complex ecological network with two interaction types. *Oikos*, 118(1):122–130, 2009.

Eliot T Miller, Damien R Farine, and Christopher H Trisos. Phylogenetic community structure metrics and null models: a review with new methods and software. *Ecography*, 2016.

Mark EJ Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.

Robert T Paine. Food web complexity and species diversity. *The American Naturalist*, 100(910):65–75, 1966.

Loïc Pellissier, Camille Albouy, Jordi Bascompte, Nina Farwig, Catherine Graham, Michel Loreau, Maria Alejandra Maglianesi, Carlos J Melián, Camille Pitteloud, Tomas Roslin, et al. Comparing species interaction networks along environmental gradients. *Biological Reviews*, 93(2):785–800, 2018.

Timothée Poisot, Daniel B Stouffer, and Sonia Kéfi. Describe, understand and predict: why do we need networks in ecology? *Functional Ecology*, 30(12):1878–1882, 2016.

Daniel B Stouffer and Jordi Bascompte. Compartmentalization increases food-web persistence. *Proceedings of the National Academy of Sciences*, 108(9):3648–3652, 2011.

Daniel B Stouffer, Marta Sales-Pardo, M Irmak Sirer, and Jordi Bascompte. Evolutionary conservation of species' roles in food webs. *Science*, 335(6075):1489–1492, 2012.

Elisa Thébault and Colin Fontaine. Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science*, 329(5993):853–856, 2010.

Jason M Tylianakis and Rebecca J Morris. Ecological networks across environmental gradients. *Annual Review of Ecology, Evolution, and Systematics*, 48, 2017.

Jason M Tylianakis, Raphael K Didham, Jordi Bascompte, and David A Wardle. Global change and species interactions in terrestrial ecosystems. *Ecology Letters*, 11(12):1351–1363, 2008.

Konstans Wells, Robert B O'Hara, Martin Pfeiffer, Maklarin B Lakim, Trevor N Petney, and Lance A Durden. Inferring host specificity and network formation through agent-based models: tick–mammal interactions in borneo. *Oecologia*, 172(2):307–316, 2013.

Peter Yodzis. Local trophodynamics and the interaction of marine mammals and fisheries in the benguela ecosystem. *Journal of Animal Ecology*, 67(4):635–658, 1998.

Part II

NETWORK METRICS

1

UNMASKING STRUCTURAL PATTERNS IN INCIDENCE MATRICES: AN APPLICATION TO ECOLOGICAL DATA

Bernat Bramon Mora[1], Giulio V. Dalla Riva[2], Daniel B. Stouffer[1]

1 Centre for Integrative Ecology, School of Biological Sciences, University of Canterbury, Christchurch, New Zealand

2 School of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand

1.1 *Abstract*

Null models have become a crucial tool for understanding structure within incidence matrices across multiple biological contexts. For example, they have been widely used for the study of ecological and biogeographic questions, testing hypotheses regarding patterns of community assembly, species co-occurrence, and biodiversity. However, to our knowledge we remain without a general and flexible approach to study the mechanisms explaining such structures. Here, we provide a method for generating "correlation-informed" null models, which combine the classic concept of null models and tools from community ecology, like joint statistical modeling. Generally, this model allows us to assess whether the information encoded within any given correlation matrix is predictive for explaining structural patterns observed within an incidence matrix. To demonstrate its utility, we apply our approach to two different case studies that represent examples of common scenarios encountered in community ecology. First, we use a phylogenetically-informed null model to detect a strong evolutionary fingerprint within empirically observed food webs, reflecting key differences in the impact of shared evolutionary history when shaping the interactions of predators or prey. Second, we use multiple informed null models to identify which factors determine structural patterns of species assemblages, focusing in on the study of nestedness and the influence of site size, isolation, species range and species richness. In addition to offering a versatile way to study the mechanisms shaping the structure of any incidence matrix, including those describing ecological communities, our approach can also be adapted further to test even more sophisticated hypotheses.

1.2 Introduction

Null models are an integral part of modern ecology and biogeography and provide a crucial statistical tool to test hypotheses regarding phenomena such as community assembly (Gotelli, 2000; Gotelli and Entsminger, 2001; Manly, 2006; Gotelli and Ulrich, 2010; Miller et al., 2016). The underlying idea behind the use of any null model is that comparisons of real data to randomly generated data can provide insights into how biological data is structured, as well as the potential mechanisms explaining such structure. Following this idea, a structural pattern found in an observed biological system is only regarded as statistically meaningful if it is not reproducible by a random model and is therefore unlikely to be found purely by chance.

In the case of ecological networks or species assemblages, null models are often based around sampling and shuffling species' interactions or presence/absence data (Gotelli, 2001), respectively. Therefore, the specific null hypothesis that is being tested with such a null model is entirely defined by the underlying randomization strategy (Gotelli and Entsminger, 2001; Ulrich and Gotelli, 2013; Rohr et al., 2014). As such, standard practice is for the randomization strategy of any null model to be generated in a way that includes some biological information while intentionally excluding other information. The differences observed between the empirical data and the data generated by the null model are then assumed to be a direct consequence of the omission of such information.

This approach, however, has not been without some controversy, since the choice of an inadequate null model may lead to artefactual conclusions (Gotelli and Graves, 1996; Manly, 2006). For example, using null models to identify the mechanisms underlying the structure of biological data can be ambiguous because there is not always a single way of introducing specific information into a model; therefore, these hypothesized mechanisms can only be supported by some evidence rather than a definitive proof. Perhaps more importantly, the randomization strategy may neglect some factors that could be responsible for a particular structural pattern. At times, this omission is due to insufficient prior evidence to support the idea that some unforeseen factor is potentially an important driver. At others, it arises due to the apparent difficulty with which to include such information into the randomization strategy. For example, null models employed to community-scale data in ecology often ignore the fact that species are part of a hierarchically structured phylogeny (Cavender-Bares et al., 2004), and thus, the idea that observed structural regularities may potentially be explained most parsimoniously as the outcome of a complex evolutionary process (Bersier and Kehrli, 2008; Gómez et al., 2010).

This present work is an attempt to overcome the aforementioned difficulties by combining the classic concept of a null model and the ideas underlying joint modeling in community ecology. Joint models are a set of statistical tools for integrating environmental predictors and species interactions into a common framework (Warton et al., 2015). These tools have been very helpful for understanding species richness and co-occurrence in ecological communities (Pollock et al., 2014; Ovaskainen et al., 2010), and we use them here to expand beyond the traditional null model approach. In particular, we present a correlation-informed null model that flexibly incorporates biologically relevant information as an ingredient for the null hypotheses as opposed to post-hoc tests of the influence of those factors on the structure of biological data or on null model comparisons (Wang et al., 2010; Matthews et al., 2015). For example, given a particular ecological community, a correlation-informed null model generates a random community that is informed by any given correlation matrix. This new approach therefore provides a methodological framework to assess the importance of any measurable species trait (e.g. phylogenetic relatedness, body size or species' tolerance to environmental conditions), habitat properties (e.g. ecosystem type, geographical distance or altitude), or combinations of these, on the structural patterns observed within such community data.

In order to demonstrate the versatility and power of the method presented here, we revisit examples from the literature that are emblematic of common problems encountered across community ecology. First, we apply the method to test whether or not a null model accounting for species' shared evolutionary history can reproduce the structural properties observed in empirical food webs. To do so, we use a phylogenetically-informed null model, which allows us to evaluate whether or not the structure of empirical and simulated food webs appears non-random when accounting for potential conservation of interactions. Second, we analyze the factors that influence the structure of species assemblages, focusing in particular on the effect of non-independence between sample sites. Using different correlation-informed null models, we unmask the factors of one of the most used patterns in island biogeography studies. Though we have chosen to frame the methodology in an ecological context, note that the correlation-informed null model can be generalized to study the structure of any system that can be represented by association data and whose components can be related by an underlying correlation structure.

1.3 Materials and methods

1.3.1 The null-model approach

Uninformed null models

The structure of many systems is commonly described using an incidence matrix. This incidence matrix A describes the relationship between two given interacting sets $\{i\}$ and $\{j\}$, where every element of the matrix A_{ij} is set to 1 when a relationship between i and j is present in the community, and 0 otherwise. For example, in ecology, a species assemblage can be conveyed by a matrix representing the presence/absence of different species across a set of sites whereas an ecological network can similarly be represented by a matrix characterizing the presence/absence of interactions between two sets of species (e.g predators and prey, plants and pollinators, or hosts and parasites). For the sake of simplicity, in the remainder of the methods, we will call any element $A_{ij} = 1$ a "link" $i \leftarrow j$ even though in a species assemblages this would not be an interaction in the standard ecological sense of the term.

With limited exceptions (e.g. Paine 1988 or Dunne et al. 2002), the statistical significance of any structural pattern in an incidence matrix is conditioned to the chosen null hypothesis (Rohr et al., 2014), which is generally described by an ensemble of randomized matrices. The vast majority of null models can follow either a probabilistic or a fixed algorithm to generate such randomized matrices (Fortuna et al., 2010). The probabilistic approach samples the matrix elements based on the total number of links of both row and column elements (Cook and Quinn, 1998; Bascompte et al., 2003), preserving approximates of their overall distributions. The fixed strategy, on the other hand, randomizes the possible links by either recursively swapping the existing ones ('swap' algorithm; Connor and Simberloff, 1979) or randomly creating them ('fill' algorithm; Sanderson et al., 1998), in such a way that they exactly match constraints imposed by row and column marginals (Gotelli and Entsminger, 2003).

The randomization strategy used here is based around the swap algorithm (Connor and Simberloff, 1979)—also referred to as fixed-fixed null model. That is, we use a Markov-chain Monte-Carlo switching algorithm to iteratively select existing links and swap them, provided that these swaps agree with the imposed constraints (Milo et al., 2003; Itzkovitz et al., 2004). For instance, for the purpose of randomizing a matrix A while preserving both the degree of row and column elements, the algorithm would repeatedly select two existing links $i \leftarrow j$ and $l \leftarrow m$ at random, and transform them into $i \leftarrow m$ and $l \leftarrow j$ on the condition that they are not already present in the community.

Importantly, the standard form of this randomization strategy swaps any pair of existing links with equal prior probability. That is, in every iteration of the randomization process, the selection criteria for the choice of the swapping links is uninformed, implying that any two links are equally likely to be shuffled as long as such shuffling agrees with the other imposed structural constraints.

Note that the swap algorithm does not establish a minimum number of iterations— also referred to as 'swap trials'—needed in order to obtain fully randomized incidence matrices; this will depend on the size and structure of the incidence matrix being randomized. Miklós and Podani (2004) recommend ensuring that the number of trials is such that the expected number of actual swaps is twice the number of 1's in the incidence matrix. For the purpose of studying structural patterns in randomized incidence matrices, however, we would suggest making sure that an increase in the number of swap trials does not lead to any changes to the average representation of such structural patterns

Correlation-informed null models

In contrast to the uninformed null model, we introduce a manner in which to "inform" the swapping algorithm so that the probability of randomizing different links depends on underlying biological information—i.e. information on additional constraints or tendencies in the natural phenomena. To do so, we modify the randomization process in order to account for the information encoded within a specified correlation matrix. Specifically, we calculate estimates of the probability to observe any given link in an incidence matrix by means of a Generalized Linear Mixed Model (Ives and Helmus, 2011; Rafferty and Ives, 2013; Pearse et al., 2014, 2015). Given a row i from an incidence matrix A of size $n \times m$ and a correlation matrix V_{col} of size $m \times m$ relating the corresponding m columns, the probability of observing a link between row i and column j can be estimated by fitting the observed links A_{ij} to the following logistic regression:

$$\text{logit}\left(p_{ij}\right) = \alpha_i + b_{ij} \tag{1}$$

where α_i is a constant intercept and b_{ij} is a Gaussian distributed random effect with mean 0 accounting for the correlation matrix (Fig. 1). The covariance matrix of b_{ij} is $\sigma_i^2 V_{col}$, which represents an estimated scalar multiplied by the $m \times m$ correlation matrix. Therefore, the estimation of the scalar σ_i roughly reveals how well the observed links can be predicted by the correlation matrix V_{col} (Ives and Helmus, 2011). As this regression is performed for every row i of the incidence matrix A, the sample size over which the parameters are inferred is exactly the number of column elements m contained in each

row. Note that this same estimation can be performed for a correlation matrix V_{row} that relates the n rows by instead fitting the model to the transpose of the incidence matrix.

The estimated probabilities p_{ij} provide then a way of weighting the randomization process based on the correlation matrix. That is, we can introduce a bias in the null model so that the swap algorithm transforms two randomly selected links $i \leftarrow j$ and $l \leftarrow m$ into $i \leftarrow m$ and $l \leftarrow j$ according to the joint probability $p_{im} \times p_{lj}$, again provided that this change agrees with any other potential constraints imposed on the model (Fig. 1). This bias to the null model defines the correlation-informed null model. Note that the estimation of the probabilities p_{ij} is done prior to the randomization process; therefore, every step of the swap algorithm is informed relative to the original incidence matrix. Also, it is worth pointing out that the chosen joint probability assumes independence of interactions, and more sophisticated approaches could also be taken into consideration when combining the estimated probabilities.

Importantly, the correlation matrix used to fit the incidence matrix can (i) either provide valuable information to explain the observed links or (ii) appear completely uninformative to them. In the former case, the estimated probabilities will present an heterogeneous pattern whereas in the latter case they will tend to show a uniform distribution. Moreover, an informative correlation matrix does not imply a more predictive null model since the information provided might be irrelevant to explain the particular pattern that is ultimately being tested.

Misinformed null models

Following the definition of the correlation-informed null model, we can also define a misinformed null model, where the randomization process is itself informed by randomized correlation matrices. That is, given an incidence matrix A and correlation matrix V_{col}, we generate every random network A^* of a misinformed null model as follows: we first randomize V_{col}, symmetrically permuting the row and column identities; we next estimate the probabilities p_{ij} using the randomized matrix V^*_{col}; and we finally use p_{ij} to "inform" the swapping algorithm as described for the correlation-informed null model. A misinformed null model is necessary because it serves as a control model for the correlation-informed counterpart. This is because it allows us to test that a null model informed with the "wrong" correlation structure—which is a form of overfitting—does not lead to artefactual conclusions. We expand on this below in the section 'Model Testing' and in the Supplementary Methods (Appendix B).

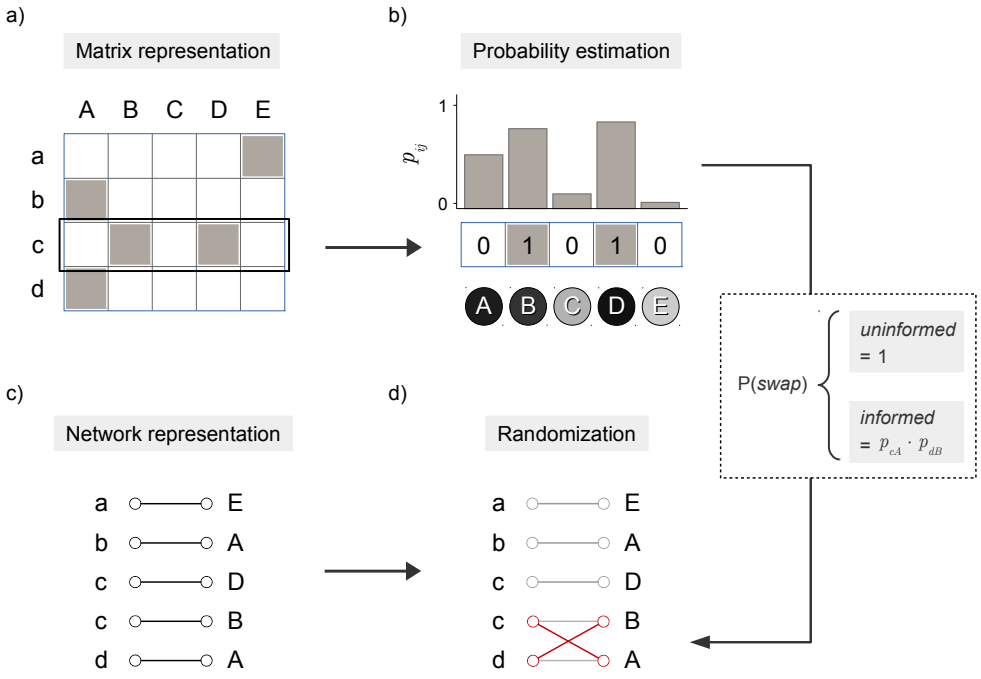


Figure 1: Graphical representation of the proposed randomization process. (a) The null model generates data by randomizing a given incidence matrix, where rows and columns represent two interacting sets $\{x\}$ and $\{Y\}$, and the grey squares indicate an interaction between elements of the sets. (b) Based on some column attribute or trait (in this example the grey tone of the circles under the graph) and the empirical matrix, we can estimate the data-informed probability of encountering any of the possible interactions of the incidence matrix. (c-d) The randomization algorithm then repeatedly swaps two randomly selected links in the network representation of (a) according to the estimated probabilities. For example, if the algorithm selected links $c \leftarrow B$ and $d \leftarrow A$, they would be swapped with probability $p = p_{cA} \cdot p_{dB}$ in the informed case. In the uninformed case, the swap would occur with probability $p = 1$.

Correlation structures

Given an incidence matrix $A_{n\times m}$, the correlation matrix $V_{col} = V_{m\times m}$ (or $V_{row} = V_{n\times n}$) defines the relationships between the m column elements (or n row elements) of A. Every element $v_{ij} = v_{ji}$ of this symmetric, positive semi-definite matrix characterizes the similarity between two columns (or rows) i and j. There are an infinite number of matrices that can be proposed as a correlation structure $V_{m\times m}$. For example, the most basic one would be a matrix such that every element v_{ij} is equal to 1, representing the case in which there are no differences across the m column elements. This basic case is important because such a correlation structure is not informative to the swap algorithm and it produces a null model that behaves exactly

as its uninformed counterpart. Alternatively, the m column elements could instead belong to different groups, and one could use these groups to define a correlation structure $V_{m\times m}$ such that $v_{ij} = 1$ if i and j belong to the same group, and 0 otherwise. This correlation structure would then inform the null model so that the randomization process is biased following such grouping.

Similar to the groups, one can generate a suitable correlation structure given any set of continuous values (or 'traits') that describe the m columns. Such a correlation structure can then take multiple forms, from a direct measure of similarity of these set of traits to other more sophisticated forms such as exponential or Gaussian structures. For instance, we could generate an exponential correlation structure $V_{m\times m}$ for a given set of column traits $\{x_m\}$ using

$$V_{col} = (1 - N)\exp\left(-\frac{D}{\max D}\right), \tag{2}$$

where D characterizes the distance matrix across all traits such that d_{ij} is the Euclidean distance between any two column traits x_i and x_j, and N is a matrix such that every element in the diagonal $n_{ii} = 0$ and any other element $n_{ij} = \eta$. The factor η characterizes the "nugget effect" for this correlation matrix, which is used as a way to avoid perfectly correlated off-diagonal elements. The computation of many well-known correlation structures can be done using functions within the R package *nlme* (Pinheiro et al., 2014); note, however, that the appropriateness of each will depend on the precise question being studied.

Quantifying over- and under-representation

To test whether or not any structural pattern observed in an empirical incidence matrix is significantly non-random compared to the data generated by a null model, we use the pattern's z-score. To understand this comparison, let's define the measure of an arbitrary structural pattern $k = k(A)$ of an adjacency matrix A. This property could characterize simple aspects of the adjacency matrix such as the total number of links or other more complex metrics of the way in which the links are distributed within the matrix. Following this, such structural pattern could also be measured in an ensemble of randomized matrices $\{A^*\}$ generated by a given null model, defining a null distribution of measures $\{k^*\}$. The pattern's z-score can then be defined as

$$z = \frac{k - \langle\{k^*\}\rangle}{\sigma_{\{k^*\}}}, \tag{3}$$

where $\langle\{k^*\}\rangle$ is the average measure of the structural pattern in the random ensemble and $\sigma_{\{k^*\}}$ is the corresponding standard deviation.

A positive z indicates that the observed pattern is overrepresented in the empirical matrix, and significantly so for values greater than 1.96. Likewise, a negative z indicates that the pattern is underrepresented, and the threshold for significance is -1.96.

1.3.2 *Applications to ecological data*

Food webs and network motifs

The first emblematic example that we revisit from the literature is the study of the evolutionary history behind food-web structure. In particular, we studied how well species' evolutionary relationships can explain observed patterns of interaction in food webs. To do so, we analyzed 10 empirical food webs from small streams of the Taieri River in New Zealand comprising fish, macroinvertebrates and algae (Townsend et al., 1998). They are taxonomically highly resolved food webs—taxonomically or trophically related species were always considered independently—and range in size from 78 to 113 species. These food webs are from habitats that present many similarities (i.e. all sites were from grassland catchments and included at least one pool and one riffle) but still differ in fundamental ways (including but not limited to different size, altitude, stream depth and land-use).

For each of the 10 food webs, we focused on the analysis of so-called food-web motifs—connected sub-graphs representing the different patterns of interactions between a subset of species (Stouffer et al., 2007). The frequency of appearance of each of these subgraphs within a network defines a structural property that has proven to be a very powerful network metric to understand food-web structure (Bramon Mora et al., 2018). When compared to a null hypothesis, this network metric has been shown to be very non-random, presenting consistent patterns of over- and under-representation (Milo et al., 2002; Stouffer et al., 2007; Stouffer and Bascompte, 2010; Baker et al., 2015; Trøjelsgaard and Olesen, 2016). We specifically focused on the study of the frequency of appearance of three-species food-web motifs, which have already been shown to be non-randomly represented in the dataset used here (Stouffer et al., 2007). To do so, we used the tools provided by the Python module 'pymfinder' (Bramon Mora et al., 2018).

Species assemblages and nestedness

As a second example, we analyzed different factors that have been shown to influence the structure of species assemblages. Specifically, we explored how well possible spatial autocorrelations or area similarity between sample sites as well as island species richness and species range similarity can explain the structural patterns observed in these

communities. To do so, we used the floristic database published by Marx et al. (2015b), which reports the distribution of 366 species of vascular plants across 80 islands from the San Juan archipelago (Marx et al., 2015a). The data were compiled between 2005 and 2010 and restricted to the smaller islands of the archipelago ($<$ 25 ha). This database also provides information on the size and geographic centroid of the islands.

In this case, we focused on the study of nestedness (Patterson and Atmar, 1986), a common measure of assemblage structure. A species-sites assemblage is said to be nested when sites with fewer species contain a subset of the species present in more abundant sites. Although there are multiple algorithms that define a measure for nestedness (Ulrich et al., 2009), we used the nestedness calculator NODF (Rodríguez-Gironés and Santamaría, 2006), which returns a value close to 100 when the community is highly nested and close to 0 otherwise.

Model testing

To validate the models before analyzing the empirical data, we benchmark tested them using artificially-generated structured and random data. We decided to use two tests that mimicked the two empirical datasets chosen to introduce the method. In particular, we first generated artificial food webs and species assemblages and informative correlation matrices for their components ('Supplementary Methods' section of Appendix B). Then, we studied the motif representation and nested patterns found in the food webs and species assemblages, respectively, comparing the performance of the uninformed, correlation-informed, and misinformed null models ('Supplementary Results' section of Appendix B). As expected, we found the uninformed and misinformed null models to showcase very similar performance—showing very similar patterns of over- and under-representation—while the correlation-informed null model was instead able to shed light on the structure of the generated data (Supplementary Figure 20). This is important because it implies that correlation structures encoding information regarding the process in which the data is generated are informative to the null model, but other unrelated correlation structures do not affect the model's performance. Finally, we performed the same tests using random data, where all the models showed the same over- and under-representation of the structural patterns (Supplementary Figure 20).

1.4 Results

1.4.1 Application to food webs

For each of the 10 empirical food webs, we first analyzed the three-species motif representation using the uninformed null model. We generated data with this model by using fixed-fixed algorithm, shuffling species' interactions while conserving each species' number of prey and predators, and the distribution of single, double and cannibal links (Itzkovitz et al., 2004; Stouffer et al., 2007). The reason for these constraints is that this type of randomization preserves the total numbers of prey and predators of all species and the two-species motif structures; therefore, it ensures that the over- or underrepresentation of a motif of size three is not due to the over- or under- representation of a particular sub-pattern (Milo et al., 2002; Artzy-Randrup et al., 2004; Itzkovitz et al., 2004). We found that three different motif structures were significantly overrepresented in all 10 networks (Fig. 2): the motifs describing a simple food chain, exploitation competition, and apparent competition. We likewise found that the motifs representing omnivory and a three-species trophic loop were consistently underrepresented in every food web.

We then performed the same analysis using the phylogenetically-informed null model. To do so, we first estimated phylogenies for the different species forming the 10 food webs under study ('Supplementary Methods' section of Appendix B) and generated the corresponding phylogenetic covariance matrices using the function 'vcv' from the R package *APE* (Paradis et al., 2004). Then, we weighted the randomization strategy used in the uninformed case to account for the information encoded within the estimated phylogenies. To achieve this, we calculated the interaction probabilities of the food webs through Eqn (1), considering the phylogenetic covariance matrices as correlation matrices. These probabilities can be estimated following two different perspectives: the predator's diet and the prey's consumers. Given any interaction $i \leftarrow j$, the former describes the probability of the predator i consuming j given the phylogenetic relationships between the prey species whereas the latter represents the probability of the prey j being consumed by i given the phylogenetic relationships between the predator species.

With these two phylogenetically-informed null models, we found the same pattern of over- and underrepresentation as that observed when using the uninformed null model (Fig. 2). In this case, however, the phylogeny appears to be particularly informative for determining food-web structure since data generated by the null model is much better at reproducing the empirical motif representation. Moreover, the results present key differences between the null model account-

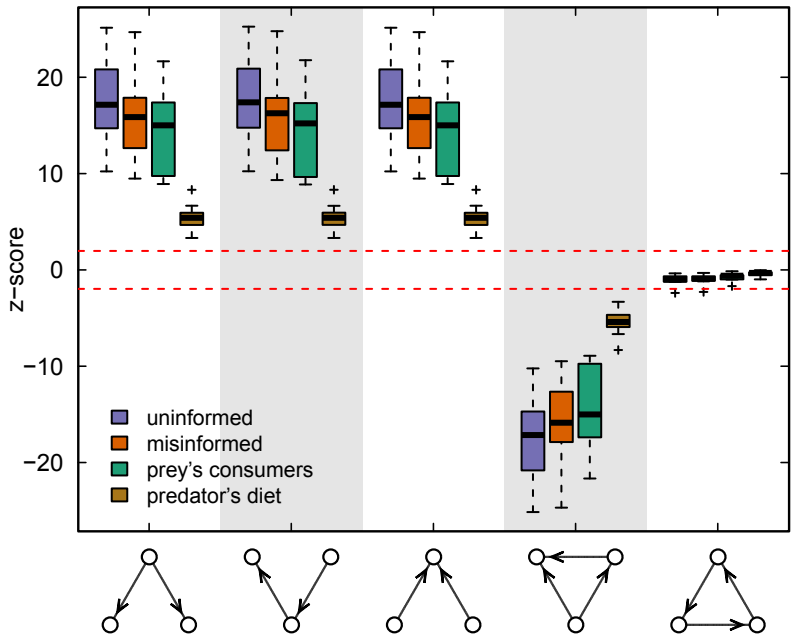


Figure 2: The effect of the phylogenetic relationships between species on the motif representation within a set of food webs. For all motifs, the arrow indicates the transfer of energy from prey to predators. The boxes contain the z-scores for each motif according to the different null models. The boxes group all food webs, extending from the lower to upper quartile values of the data, with a line at the median. The color of the boxes indicates the null model used: an uninformed null model (uninformed), a misinformed null model (misinformed), a null model accounting for the phylogenetic relationships in preys' consumers (prey's consumers), and a null model accounting for the phylogenetic relationships in predators' diets (predator's diet). The dashed red line indicates the thresholds for significance $z \leqslant -1.96$ and $z \geqslant 1.96$.

ing for the phylogenetic relationships of predators' diet and the one accounting for the phylogenetic relationships of preys' consumers. Specifically, the motif profile is best preserved when we considered the predators's diet perspective but is significantly less informative when the prey's consumers perspective is adopted (Fig. 2). Importantly, the observed differences between the two informed null models were true even when controlling for the degree of overlap between the empirical food-webs and their randomized counterparts ('Supplementary Methods' and 'Supplementary Results' sections of Appendix B). That is, such differences were not due to the number of shared links between the empirical and random structures but instead arose from the intrinsic properties of the adopted null hypotheses ('Supplementary Results' of Appendix B).

1.4.2 *Application to species assemblages*

For the species assemblage data, we first analyzed the nestedness pattern using the uninformed null model. We again followed the fixed-fixed algorithm, which is one of the most widely used approaches in biogeographic studies whereby the incidence matrix is randomized fixing both the number of species per site and the relative frequency of appearance of each species (Miklós and Podani, 2004; Ulrich and Gotelli, 2007; Almeida-Neto and Ulrich, 2011; Strona and Fattorini, 2014). We observed that this species assemblage is more nested than expected by chance, presenting a significantly high z-score (Fig. 3).

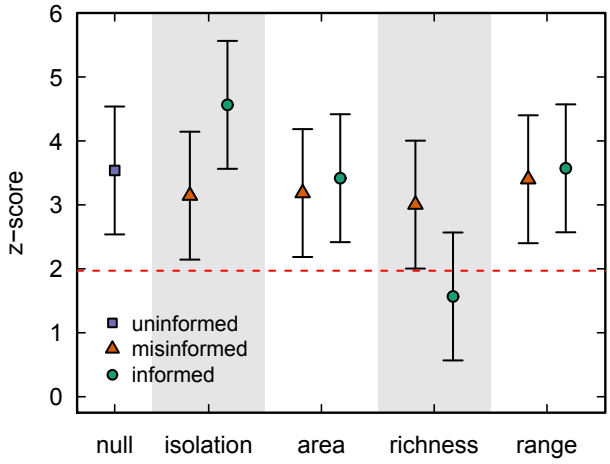


Figure 3: The effect of spatial autocorrelation, island area, island species richness and species range similarity on measures of community structure. We show the z-scores for the nestedness pattern in the distribution of vascular plants across islands from the San Juan archipelago. All plots show the results obtained using an uninformed null model (null), an isolation-informed null model (isolation), an area-informed null model (area), the richness-informed null model (richness), and the range-informed null model (range). The red dotted line indicates the threshold for significance $z \geqslant 1.96$.

Then, we used different informed null models to quantify the influence of the island isolation, island area, island species richness, and species range on the structure of this community. To do so, we first computed separate correlation structures for each of these factors. In particular, we assumed an exponential correlation by means of Eqn (2), which is a widely used approach to account for spatial autocorrelation in biogeographic studies (Dormann et al., 2007). For these particular examples, we used a nugget effect $\eta = 0.01$ to generate the correlation structure. Following this, we weighted the uninformed randomization process to account for the different correlation matrices using Eqn (1), as described in the Methods section.

The isolation-informed and area-informed null models, on one hand, showed the species assemblage to be significantly nested, presenting the same overall conclusion as the uninformed null model. That is, spatial autocorrelation and size similarity between islands in this database is not a significant predictor of the observed nested pattern (Fig. 3). On the other hand, the results obtained using the richness-informed and range-informed null models showed that while the difference in the relative frequency of appearance of each species is not a significant predictor of the observed nested pattern, the difference in the number of species per site is (Fig. 3). That is, the random matrices generated by the null model informed using the species range appeared significantly less nested than the empirical matrix; however, the random matrices generated by the null model informed using the island species richness appeared as nested as the empirical matrix (Fig. 3). In all cases, the results were also compared to the ones produced by misinformed null models, finding no apparent differences with the uninformed counterpart for this pattern.

1.5 *Discussion*

An extensive literature has been published about null models in ecology and biogeography (Gotelli and Graves, 1996; Gotelli, 2000; Miller et al., 2016), including models accounting for within-species spatial patterns (Roxburgh and Chesson, 1998; Roxburgh and Matsuki, 1999). Inspired by this work, we present here a general and flexible approach to study the mechanisms explaining the structure of biological communities. In particular, we combine the classic concept of a null model and the ideas underlying joint modeling to define a correlation-informed null model. This model allows us to assess how informative the information encoded within any given correlation matrix is for explaining the structural patterns observed within any incidence matrix. Using this approach, we focused on the study of the biological mechanisms shaping the structure of ecological networks and species assemblages. Specifically, we found (1) a strong phylogenetic component underlying food-web motifs and (2) a nested pattern in species assemblages that seems to be predominantly explained by island species richness.

In the first application of the correlation-informed null model, we studied the phylogenetic signal behind species' interactions. This idea was based on the long held assumption and frequent observation that these interactions are evolutionarily conserved (Bersier and Kehrli, 2008; Rezende et al., 2009; Gómez et al., 2010). In particular, we compared uninformed, misinformed and phylogenetically-informed null models to study the motif representation of empirical food webs. This comparison showed that the network's motif profile is largely pre-

served in data generated by a null model accounting for the phylogenetic relationships in predators' diets. In contrast, we found that this model is significantly less informative when the analogous prey's consumers perspective is adopted. First, this observation showcases how biological mechanisms can be untangled using our approach. In particular, it supports the idea of a stronger phylogenetic signal in prey range for predators than in predator range for prey (Naisbit et al., 2012) as well as a prey-selection mechanism shaping the structure of food webs (Stouffer et al., 2007). Importantly, although the effect of the phylogenetic information reveals itself as crucial to explain who interacts with whom in a food web, our results also highlight the fact that this is clearly insufficient to fully predict motif representation in prey-predator relationships.

In the second application of the correlation-informed null model, we analyzed the patterns observed in species distributions across different habitats. We focused on the study of nestedness, which is a common measure employed in biogeographic studies. Nestedness has been associated with habitat variables such as area (Watling and Donnelly, 2006; Wang et al., 2010, 2013), isolation (Kadmon, 1995; Wang et al., 2010) or land quality (Triantis and Bhagwat, 2011). Somewhat surprisingly, we found that neither isolation nor area differences between islands appear to provide particularly relevant information to explain the nested pattern observed in the distribution of vascular plants across islands from the San Juan archipelago. One potential explanation for this lack of predictive power could be the fact that the biggest islands of the archipelago were excluded from the analysis (Marx et al., 2015b). This notwithstanding, we found that island species richness can instead explain the nested pattern. This observation is important because it suggests that nestedness is perhaps little more than an artifact of island species richness that becomes tautological when one controls for it. Moreover, the observation is in contrast to the results found when using the range-inform the null models, finding that species range is instead a poor predictor of the nested pattern observed in the species assemblage.

As a key step moving forward, it could be worth adapting the strategies presented in this work to inform other models from neutral theory of island biogeography, expanding the framework to new randomization strategies. In addition, the scenarios presented here provide only an introduction of the possible applications for any such correlation-informed null model. For example, one could also focus on the structure of ecological bipartite networks (e.g. plant-pollinator, host-parasitoid, seed-dispersal, etc.) and the drivers determining observed non-random patterns (e.g. modularity, uniqueness, centrality, etc.). We could evaluate whether or not there is a dominant trait from

a particular group shaping the interactions of those networks—e.g. nectar depths of plants or proboscis length of pollinators in plant-pollinator networks (Stang et al., 2009) and seed or beak size in seed-dispersal networks (Dehling et al., 2014). Alternatively, we could examine the structural differences observed between different network types and assess which are the factors explaining such differences—e.g. comparing pollination and herbivory network architectures accounting for the evolutionary relationships of all their constituents (Thébault and Fontaine, 2010).

Here, we have sought to showcase some of the basic applications of the correlation-informed null model, but there are many other questions that could be addressed using the same approach. For instance, the examples presented here have only considered the effect of a single correlation matrix; however, one could take into account higher order correlations or simultaneously consider multiple correlation matrices to inform the same null model. Indeed, we could consider multiple random effects in Eqn (1) or combine the probabilities generated using multiple correlation matrices independently (Ives and Helmus, 2011). In an ecological context, this informed null model could study species assemblages by combining different habitat properties (e.g. soil characteristics, vegetation type, etc.) with multiple species traits (e.g. body size, phylogenetic relationships, etc.) into a generalized island biogeography study. Consequently, our correlation-informed null model offers a versatile way to study the mechanisms shaping the structure within biological data that can easily be adapted further to test even more sophisticated hypotheses. Perhaps more importantly, there are multiple systems and structural patterns outside the ecological realm for which a correlation-informed null model could be useful. Indeed, our model only requires a system whose structure can be represented as an incidence matrix. Therefore, similar analyses could be performed for systems such as protein-protein interaction networks, neuronal networks, or transcriptional regulation networks, among many others.

1.6 *Data accessibility*

The method presented here is implemented as an R package (R Core Team, 2014) named *resoldre* and is available online at `https://github.com/bernibra/RESOLDRE`. The key components of *resoldre* call functions from the Rcpp (Eddelbuettel et al., 2011) and RcppArmadillo (Eddelbuettel and Sanderson, 2014) packages in order to speed up its use. The main function of this package is 'randomize', which can perform different informed randomizations of a given incidence matrix based on a specified correlation matrix relating its rows

or columns. The data and a description of data files are also included in the same repository.

BIBLIOGRAPHY

Mário Almeida-Neto and Werner Ulrich. A straightforward computational approach for measuring nestedness using quantitative matrices. *Environmental Modelling and Software*, 26(2):173–178, 2011.

Yael Artzy-Randrup, Sarel J Fleishman, Nir Ben-Tal, and Lewi Stone. Comment on "network motifs: simple building blocks of complex networks" and "superfamilies of evolved and designed networks". *Science*, 305(5687):1107–1107, 2004.

Nick J Baker, Riikka Kaartinen, Tomas Roslin, and Daniel B Stouffer. Species' roles in food webs show fidelity across a highly variable oak forest. *Ecography*, 38(2):130–139, 2015.

Jordi Bascompte, Pedro Jordano, Carlos J Melián, and Jens M Olesen. The nested assembly of plant–animal mutualistic networks. *Proceedings of the National Academy of Sciences*, 100(16):9383–9387, 2003.

Louis-Félix Bersier and Patrik Kehrli. The signature of phylogenetic constraints on food-web structure. *Ecological Complexity*, 5(2):132–139, 2008.

Bernat Bramon Mora, Alyssa R Cirtwill, and Daniel B Stouffer. pymfinder: a tool for the motif analysis of binary and quantitative complex networks. *bioRxiv*, page 364703, 2018.

Jeannine Cavender-Bares, Kaoru Kitajima, and Fakhri A Bazzaz. Multiple trait associations in relation to habitat differentiation among 17 floridian oak species. *Ecological Monographs*, 74(4):635–662, 2004.

Edward F Connor and Daniel Simberloff. The assembly of species communities: chance or competition? *Ecology*, 60(6):1132–1140, 1979.

Rosamonde R Cook and James F Quinn. An evaluation of randomization models for nested species subsets analysis. *Oecologia*, 113 (4):584–592, 1998.

D Matthias Dehling, Till Töpfer, H Martin Schaefer, Pedro Jordano, Katrin Böhning-Gaese, and Matthias Schleuning. Functional relationships beyond species richness patterns: trait matching in plant–bird mutualisms across scales. *Global Ecology and Biogeography*, 23 (10):1085–1093, 2014.

Carsten F. Dormann, Jana M McPherson, Miguel B. Araujo, Roger Bivand, Janine Bolliger, Gudrun Carl, Richard G. Davies, Alexandre Hirzel, Walter Jetz, W. Daniel Kissling, Ingolf Kuehn, Ralf Ohlemueller, Pedro R. Peres-Neto, Bjoern Reineking, Boris Schroeder, Frank M. Schurr, and Robert Wilson. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30(5):609–628, 2007.

Jennifer A Dunne, Richard J Williams, and Neo D Martinez. Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecology Letters*, 5(4):558–567, 2002.

Dirk Eddelbuettel and Conrad Sanderson. Rcpparmadillo: accelerating R with high-performance C++ linear algebra. *Computational Statistics and Data Analysis*, 71:1054–1063, 2014.

Dirk Eddelbuettel, Romain François, J Allaire, John Chambers, Douglas Bates, and Kevin Ushey. Rcpp: seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18, 2011.

Miguel A Fortuna, Daniel B Stouffer, Jens M Olesen, Pedro Jordano, David Mouillot, Boris R Krasnov, Robert Poulin, and Jordi Bascompte. Nestedness versus modularity in ecological networks: two sides of the same coin? *Journal of Animal Ecology*, 79(4):811–817, 2010.

José M Gómez, Miguel Verdú, and Francisco Perfectti. Ecological interactions are evolutionarily conserved across the entire tree of life. *Nature*, 465(7300):918–921, 2010.

Nicholas J Gotelli. Null model analysis of species co-occurrence patterns. *Ecology*, 81(9):2606–2621, 2000.

Nicholas J Gotelli. Research frontiers in null model analysis. *Global Ecology and Biogeography*, 10(4):337–343, 2001.

Nicholas J Gotelli and Gary L Entsminger. Swap and fill algorithms in null model analysis: rethinking the knight's tour. *Oecologia*, 129 (2):281–291, 2001.

Nicholas J Gotelli and Gary L Entsminger. Swap algorithms in null model analysis. *Ecology*, 84(2):532–535, 2003.

Nicholas J Gotelli and Gary R Graves. *Null models in ecology.* Smithsonian Institution Press, Washington, D.C., USA., 1996.

Nicholas J Gotelli and Werner Ulrich. The empirical Bayes approach as a tool to identify non-random species associations. *Oecologia*, 162(2):463–477, 2010.

S Itzkovitz, R Milo, N Kashtan, MEJ Newman, and U Alon. Reply to "comment on 'subgraphs in random networks' ". *Physical Review E*, 70(5):058102, 2004.

Anthony R Ives and Matthew R Helmus. Generalized linear mixed models for phylogenetic analyses of community structure. *Ecological Monographs*, 81(3):511–525, 2011.

Ronen Kadmon. Nested species subsets and geographic isolation: a case study. *Ecology*, 76(2):458–465, 1995.

Bryan FJ Manly. *Randomization, bootstrap and Monte Carlo methods in biology*, volume 70. CRC Press, 2006.

Hannah E Marx, David E Giblin, Peter W Dunwiddie, and David C Tank. Data from: Deconstructing darwin's naturalization conundrum in the san juan islands using community phylogenetics and functional traits, 2015a. URL http://dx.doi.org/10.5061/dryad.m88g7.

Hannah E Marx, David E Giblin, Peter W Dunwiddie, and David C Tank. Deconstructing darwin's naturalization conundrum in the san juan islands using community phylogenetics and functional traits. *Diversity and Distributions*, 2015b.

Thomas J Matthews, H Eden W Cottee-Jones, and Robert J Whittaker. Quantifying and interpreting nestedness in habitat islands: a synthetic analysis of multiple datasets. *Diversity and Distributions*, 21 (4):392–404, 2015.

István Miklós and János Podani. Randomization of presence–absence matrices: comments and new algorithms. *Ecology*, 85(1):86–92, 2004.

Eliot T Miller, Damien R Farine, and Christopher H Trisos. Phylogenetic community structure metrics and null models: a review with new methods and software. *Ecography*, 2016.

Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.

Ron Milo, Nadav Kashtan, Shalev Itzkovitz, Mark EJ Newman, and Uri Alon. On the uniform generation of random graphs with prescribed degree sequences. *arXiv preprint cond-mat/o312028*, 2003.

Russell E Naisbit, Rudolf P Rohr, Axel G Rossberg, Patrik Kehrli, and Louis-Félix Bersier. Phylogeny versus body size as determinants of food web structure. *Proceedings of the Royal Society B: Biological Sciences*, 279(1741):3291–3297, 2012.

Otso Ovaskainen, Jenni Hottola, and Juha Siitonen. Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. *Ecology*, 91(9):2514–2521, 2010.

Robert T Paine. Food webs: road maps of interactions or grist for theoretical development? *Ecology*, 69(6):1648–1654, 1988.

Emmanuel Paradis, Julien Claude, and Korbinian Strimmer. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20(2):289–290, 2004.

Bruce D Patterson and Wirt Atmar. Nested subsets and the structure of insular mammalian faunas and archipelagos. *Biological Journal of the Linnean Society*, 28(1-2):65–82, 1986.

William D Pearse, Andy Purvis, Jeannine Cavender-Bares, and Matthew R Helmus. Metrics and models of community phylogenetics. In *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology*, pages 451–464. Springer, 2014.

William D Pearse, Marc W Cadotte, Jeannine Cavender-Bares, Anthony R Ives, Caroline M Tucker, Steve C Walker, and Matthew R Helmus. pez: phylogenetics for the environmental sciences. *Bioinformatics*, 31(17):2888–2890, 2015.

José Pinheiro, D Bates, S DebRoy, D Sarkar, and the R Core Team. nlme: linear and nonlinear mixed effects models. 2014. URL http://CRAN.R-project.org/package=nlme. R package version 3.1-117.

Laura J Pollock, Reid Tingley, William K Morris, Nick Golding, Robert B O'Hara, Kirsten M Parris, Peter A Vesk, and Michael A McCarthy. Understanding co-occurrence by modelling species simultaneously with a joint species distribution model (jsdm). *Methods in Ecology and Evolution*, 5(5):397–406, 2014.

R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2014. URL http://www.R-project.org.

Nicole E Rafferty and Anthony R Ives. Phylogenetic trait-based analyses of ecological networks. *Ecology*, 94(10):2321–2333, 2013.

Enrico L Rezende, Eva M Albert, Miguel A Fortuna, and Jordi Bascompte. Compartments in a marine food web associated with phylogeny, body mass, and habitat structure. *Ecology Letters*, 12(8): 779–788, 2009.

Miguel A Rodríguez-Gironés and Luis Santamaría. A new algorithm to calculate the nestedness temperature of presence–absence matrices. *Journal of Biogeography*, 33(5):924–935, 2006.

Rudolf P Rohr, Serguei Saavedra, and Jordi Bascompte. On the structural stability of mutualistic systems. *Science*, 345(6195):1253497, 2014.

Stephen H Roxburgh and Peter Chesson. A new method for detecting species associations with spatially autocorrelated data. *Ecology*, 79 (6):2180–2192, 1998.

Stephen H Roxburgh and Mamoru Matsuki. The statistical validation of null models used in spatial association analyses. *Oikos*, 85(1): 68–78, 1999.

James G Sanderson, Michael P Moulton, and Ralph G Selfridge. Null matrices and the analysis of species co-occurrences. *Oecologia*, 116 (1-2):275–283, 1998.

Martina Stang, Peter GL Klinkhamer, Nickolas M Waser, Ingo Stang, and Eddy van der Meijden. Size-specific interaction patterns and size matching in a plant–pollinator interaction web. *Annals of Botany*, 103(9):1459–1469, 2009.

Daniel B Stouffer and Jordi Bascompte. Understanding food-web persistence from local to global scales. *Ecology Letters*, 13(2):154–161, 2010.

Daniel B Stouffer, Juan Camacho, Wenxin Jiang, and Luís A Nunes Amaral. Evidence for the existence of a robust pattern of prey selection in food webs. *Proceedings of the Royal Society B: Biological Sciences*, 274(1621):1931–1940, 2007.

Giovanni Strona and Simone Fattorini. On the methods to assess significance in nestedness analyses. *Theory in Biosciences*, 133(3-4): 179–186, 2014.

Elisa Thébault and Colin Fontaine. Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science*, 329(5993):853–856, 2010.

Townsend, Thompson, McIntosh, Kilroy, Edwards, and Scarsbrook. Disturbance, resource supply, and food-web architecture in streams. *Ecology Letters*, 1(3):200–209, 1998.

Kostas A Triantis and Shonil A Bhagwat. Applied island biogeography. *Conservation Biogeography*, pages 190–223, 2011.

Kristian Trøjelsgaard and Jens M Olesen. Ecological networks in motion: micro-and macroscopic variability across scales. *Functional Ecology*, 2016.

Werner Ulrich and Nicholas J Gotelli. Null model analysis of species nestedness patterns. *Ecology*, 88(7):1824–1831, 2007.

Werner Ulrich and Nicholas J Gotelli. Pattern detection in null model analysis. *Oikos*, 122(1):2–18, 2013.

Werner Ulrich, Mário Almeida-Neto, and Nicholas J Gotelli. A consumer's guide to nestedness analysis. *Oikos*, 118(1):3–17, 2009.

Yanping Wang, Yixin Bao, Mingjian Yu, Gaofu Xu, and Ping Ding. Nestedness for different reasons: the distributions of birds, lizards and small mammals on islands of an inundated lake. *Diversity and Distributions*, 16(5):862–873, 2010.

Yanping Wang, Ping Ding, Shuihua Chen, and Guangmei Zheng. Nestedness of bird assemblages on urban woodlots: implications for conservation. *Landscape and Urban Planning*, 111:59–67, 2013.

David I Warton, F Guillaume Blanchet, Robert B O'Hara, Otso Ovaskainen, Sara Taskinen, Steven C Walker, and Francis KC Hui. So many variables: joint modeling in community ecology. *Trends in Ecology and Evolution*, 30(12):766–779, 2015.

James I Watling and Maureen A Donnelly. Fragments as islands: a synthesis of faunal responses to habitat patchiness. *Conservation Biology*, 20(4):1016–1025, 2006.

Part III

STRUCTURE

2

IDENTIFYING A COMMON BACKBONE OF INTERACTIONS UNDERLYING FOOD WEBS FROM DIFFERENT ECOSYSTEMS

Bernat Bramon Mora[1], Dominique Gravel[2,3], Luis J. Gilarranz[4], Timothée Poisot[2,5,6], Daniel B. Stouffer[1]

1 Centre for Integrative Ecology, School of Biological Sciences, University of Canterbury, Christchurch, New Zealand.

2 Québec Centre for Biodiversity Sciences, McGill University, Montréal, Canada.

3 Canada Research Chair on Integrative ecology, Départment de Biologie, Université de Sherbrooke, Sherbrooke, Canada.

4 Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland.

5 Département de Sciences Biologiques, Université de Montréal, Montréal, Canada.

2.1 *Abstract*

Although the structure of empirical food webs can differ between ecosystems, there is growing evidence of multiple ways in which they also exhibit common topological properties. To reconcile these contrasting observations, we postulate the existence of a backbone of interactions underlying all ecological networks—a common substructure within every network comprised of species playing similar ecological roles—and a periphery of species whose idiosyncrasies help explain the differences between networks. To test this conjecture, we introduce a new approach to investigate the structural similarity of 411 food webs from multiple environments and biomes. We first find significant differences in the way species in different ecosystems interact with each other. Despite these differences, we then show that there is compelling evidence of a common backbone of interactions underpinning all food webs. We expect that identifying a backbone of interactions will shed light on the rules driving assembly of different ecological communities.

2.2 *Introduction*

The structure of ecological networks—the way interactions are distributed among consumers and resources—has been shown to vary in space and time (Baiser et al., 2012; Poisot et al., 2015). Known drivers of this variation are that species composition is affected by environmental conditions, dispersal limitations, and historical contingencies (Currie et al., 2004; Thompson and Townsend, 2005; Petchey et al., 2010). Ecological interactions also vary over time and from one location to another in accordance with local changes in species abundances and traits (Poisot et al., 2015), as well as due to other intrinsic processes producing ongoing extinctions in the absence of perturbations (Allhoff et al., 2015). The nature of environmental variability in different habitats might also shape ecological networks in different ways. For instance, communities experiencing high seasonality, such as stream and lake food webs, present a strong latitudinal gradient in the number of prey and predators per species (Cirtwill et al., 2015). Moreover, the effects of disturbances like invasive species and habitat fragmentation can introduce additional variability that can also lead to changes in network structure (Morris, 2010; Aizen et al., 2008, 2012). Differences in the sampling methods can also lead to changes in the data collected, with some techniques making it hard to observe weak links in particular (Goldwasser and Roughgarden, 1997).

Despite this observed variability, many types of ecological networks also showcase a variety of common structural properties across environments (Cirtwill et al., 2015; Dunne et al., 2002a; Williams et al., 2002). For food webs, examples include relatively short food-chains (Lawton, 1989) and a roughly constant fraction of top, intermediate, and basal species (Hall and Raffaelli, 1993). The observation of these common structural properties might suggest the existence of general rules driving or constraining the assembly of all ecological communities (Piechnik et al., 2008; Holt, 2010; Williams, 2010; Gravel et al., 2011b; Cirtwill and Stouffer, 2015). Some such rules are thought to be the result of energetic or metabolic constraints in the way individual organisms process energy and materials, which could translate into some of the scaling relations observed across ecosystems (Hatton et al., 2015; Brown et al., 2004). Aspects of network structure have also been linked to ecosystems' robustness to species extinctions (Dunne et al., 2002b; Saavedra et al., 2011), persistence (Stouffer and Bascompte, 2010, 2011; Gravel et al., 2011a) and dynamical stability (Thébault and Fontaine, 2010; Allesina and Tang, 2012), which has led to some arguing that stability and feasibility are additional constraints shaping these ecological communities (Borrelli et al., 2015).

Notably, the aforementioned structural variability and commonality observed across environments need not be incompatible, though they

are often treated as such (Blüthgen, 2010). Indeed, one heretofore unexplored idea that could reconcile these two perspectives is the existence of a common 'backbone of interactions' underlying all ecological communities. Conceptually, this backbone would constitute a set of connected species within every network that play similar ecological roles and that also interact with each other in a similar manner. Extrinsic and intrinsic differences, like environmental variability or variation between local species pools, would then introduce idiosyncrasies in realized community assembly and add noise to and around the backbone.

While a backbone of interactions shared across disparate food webs might be a compelling idea, current methods for comparing network structure across communities lack the power to identify such a level of organization. In particular, existing methods are generally based around the comparison of a library of different descriptors of network structure (Williams and Martinez, 2000; Dunne et al., 2002b; Stouffer et al., 2007, 2011); however, these descriptors are summary statistics at the network level and mostly overlook the actual way ecological interactions are distributed within a network. Alternatively, one potential way to identify a backbone is by directly aligning networks in such a way as to pair up species from the different communities that play similar ecological roles (Fig. 4*A*). Doing this network alignment across a large enough dataset, the backbone of interactions could emerge as a substructure that is consistently aligned across environments.

Recent advances in network science have provided multiple methods for aligning complex networks (Flannick et al., 2006; Singh et al., 2007; Kuchaiev et al., 2010; Patro and Kingsford, 2012; Poisot et al., 2012). Most of these methods, however, focus on aligning undirected networks, making them ill-suited for ecological networks like predator-prey food webs in which the direction of interactions is particularly relevant (Allesina et al., 2015). In this study, we develop a new alignment technique specifically designed for directed networks, and we then use it to test whether or not there is a backbone of interactions across food webs. For this test, we align a collection of over 400 food webs that were compiled from multiple ecosystems—including different types of freshwater, marine and terrestrial ecosystems. For every pair of food webs, our method matches their constituent species based on their role similarity, which measures how similarly any two species are embedded within their respective communities. In particular, our method provides us with two key pieces of information. The first is a metric describing the "quality" of the alignment between food webs, which represents an overall measure of how similar two networks are to each other. The second consists of a list of the corre-

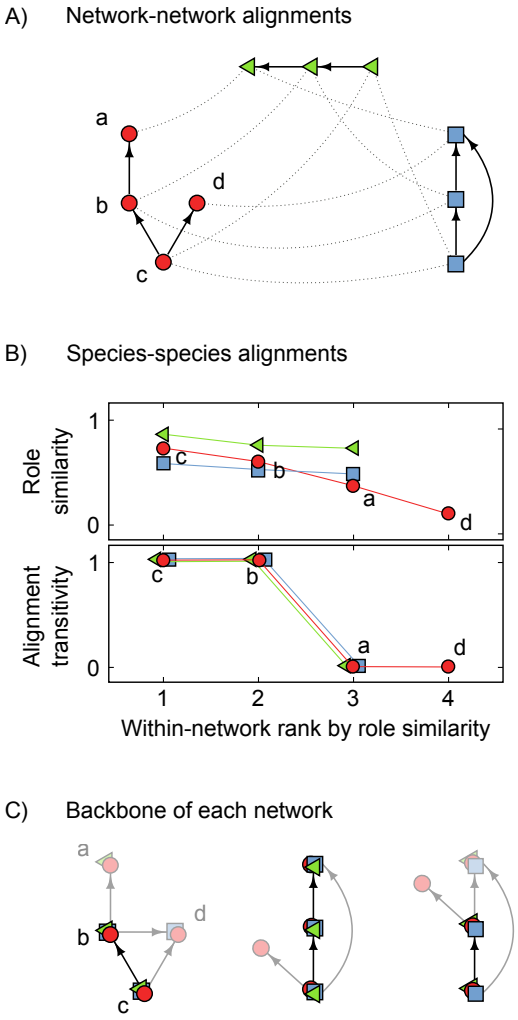


Figure 4: Network alignment and identifying a backbone of interactions. (A) An example of the optimal alignments between three simple networks. The red circles, green triangles and blue squares represent the species in each network, and the arrows indicate the direction of energy flow between those species. The dotted lines characterize the pairings of species in the three alignments between networks. (B) Given the alignments in (A), we rank species according to the average role similarity that they present across their pairings. The top panel shows the actual average role similarity, and the bottom presents the alignment transitivity of those same species. The best-aligned species from the red network is species *c*, whereas the worst aligned is species *d*. The species in the blue and green networks to which species *c* is paired are also paired, which implies that the alignment transitivity of *c* is 1. In contrast, the alignment transitivity for species *d* is 0 because there are no paired species in the blue and green networks to which *d* is paired. (C) Given the alignments in (A), we can also identify the backbone of interactions for each network. Here, the dark links are those that present the maximum overlap across network alignments and therefore characterize the backbone of interactions. The lighter links represent the periphery of such backbone.

sponding species-species pairings between those food webs, specifying the actual mapping of the alignment between them.

Here, we use the alignment quality as a metric with which to test for structural differences across ecosystems. In particular, we find that food webs from different ecosystems present significantly different network structures. We then leverage the lists of species-species pairings to identify subsets of species within every food web that align better than the rest, since these species could well constitute a backbone (Fig. 4*B*). Next, we test whether or not these subsets of species are actually linked together, and we observe that they do indeed form a connected backbone of interactions (Fig. 4*B*). To determine what these connected backbones actually look like, we finally explore the overlap of the backbones between all aligned networks to reveal the hidden structures that underly our dataset (Fig. 4*C*).

2.3 *Results*

2.3.1 *Structural differences across ecosystems*

We first analyzed the overall differences across all food webs in order to test whether or not there are significant structural differences across ecosystem types. To do so, we identified optimal alignments between every pair of food webs in our dataset, where each alignment pairs up species with similar interaction patterns in their respective networks (Methods). For each pair of food webs, we started with a random alignment and then used a simulated-annealing algorithm to progressively minimize an alignment cost function that decreases when both paired species and those species' neighbors play similar ecological roles (Methods; "Alignment algorithm" and "Algorithm tests" sections of Appendix C; Supplementary Figs 30–33).

From these pairwise alignments between all food webs in our dataset, we constructed a food-web dissimilarity matrix $\hat{E}$, where every element $\hat{e}_{ij}$ represents the 'alignment quality' between any two webs i and j (Methods). Using this matrix, we tested whether or not the alignments between food webs from the same type of ecosystem tend to be better than the ones between food webs from different ecosystem types. We found that there are indeed significant differences in the quality of the alignments between the different ecosystems (PERMANOVA; $F_{4,411} = 22.81$, $p < 0.01$; Methods). In general, this is true regardless of the choice of alignment quality metric or constraining our dataset to avoid comparing food webs with very different sizes (Fig. 5 and Supplementary Table 2; "Structural differences across ecosystems" section of Appendix C). We repeated the tests separately for every pair of ecosystem types in our dataset, finding that the ma-

jority of pairwise comparisons reinforced the idea of structural divergence between ecosystems (Supplementary Table 3 and Supplementary Fig. 21; "Pairwise comparisons between ecosystems" section of Appendix C). Based on those comparisons, the structure of freshwater stream food webs seems to be the most different when compared to all other ecosystem types.

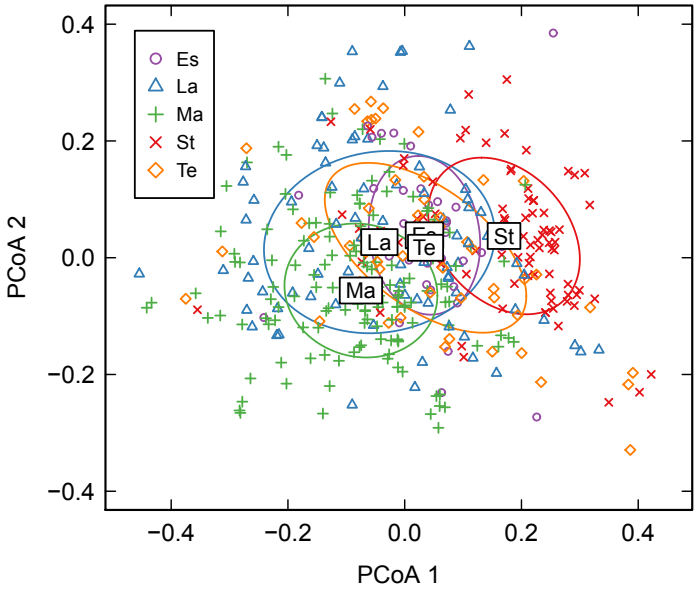


Figure 5: Principal coordinate analysis of the dissimilarity matrix $\hat{E}$ containing the normalized pairwise distances between all food webs. Each different color represents the group of networks from estuaries (Es), lakes (La), marine (Ma), streams (St) and terrestrial (Te) ecosystems. The ellipses characterize the 1 standard deviation ellipses about the group medians.

2.3.2 *Identifying backbones of interactions across food webs*

We next studied the way that individual species from different food webs were matched to each other by collectively analyzing every species-species pairing across network alignments. For every network, we ranked its species based on their average role similarity; that is, based on the average similarity between their role and the role of the species to which they were matched (Methods). Within these rankings, species that match very well—because they have very similar structural roles—will be ranked first whereas those that present a lower role similarity in their matchings will be ranked last (Fig. 4*B*). We observed that species' average role similarity can vary considerably (Fig. 6), with some species tending to align substantially better than others. Importantly, this result is independent from the ecosystem type of the food webs. That is, a ranking made based solely on the alignments of food webs within one ecosystem type is generally

very similar to a ranking based solely on alignments across different ecosystems (Fig. 6). Though we previously observed significant differences between ecosystems based on their overall network alignments, the similarity of these species-level pairings implies that the best-aligned species from a given food web will, in general, be the same for any of that web's alignments. Moreover, these species do not exclusively come from a specific trophic level, despite the fact that some trophic levels are vastly overrepresented in our data relative to others (Supplementary Fig. 22). They do, however, tend to be those with the greatest total number of interactions (Supplementary Fig. 22).

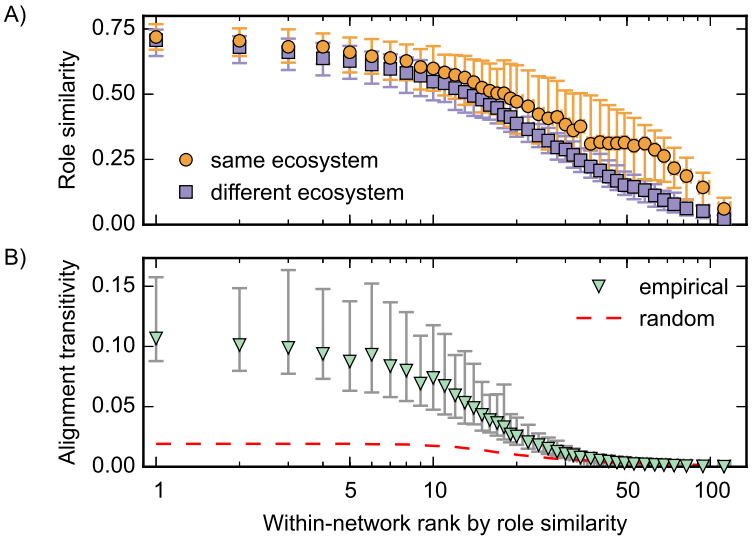


Figure 6: Ranking of species from our dataset of 411 food webs based on the average similarity between their role and the roles of the species to which they are paired across all 84255 alignments. The top panel (A) shows the observed role similarity for all species when compared to food webs from either the same (circles) or different (squares) ecosystem types. The bottom panel (B) shows the alignment transitivity observed for all species across all food webs. The red dotted line represents the expected alignment transitivity for shuffled alignments, where the number of pairings per alignment was maintained. In both panels, every point indicates the median across at least 250 species with the exception of the last point which is the median across 30 species, and the error bars characterize the interquartile range.

Our observation that every network has a set of species that align much better than the rest could be indicative of the existence of a backbone of interactions underlying all these communities. However, this observation is still not a sufficient condition for the backbone to exist. Instead, we identified two necessary conditions for the presence of a backbone of interactions: (i) the best aligned species from all networks should tend to be paired to each other; and (ii) they should also

form a connected component in their own network. To test the first condition, we studied the transitivity of species' alignments, which is a measure of how coherent a species' pairings are across alignments (Fig. 1*B*; Methods). We observed that the best aligned species show a significantly higher alignment transitivity than would be expected at random (Fig. 6). This implies that the best aligned species for the different food webs are in fact paired with each other more often than expected by chance, satisfying condition (i). Next, we indirectly tested the second condition by studying the path likelihood between species, which is a measure of how connected a set of species is within a network (Methods). For every network, we compared the subweb formed by the set of best aligned species to structures formed by equally sized random subsets of species. We found that the best aligned species tend to present a high path probability (Supplementary Fig. 23; "Connectance and path likelihood" section of Appendix C), which implies that those species are also more connected and likely to form a connected component than expected by chance, satisfying condition (ii).

In satisfying these two conditions, the evidence reveals that there likely is an underlying backbone of interactions across all the food webs in our study. However, these tests do not provide information regarding the shape of such a backbone. To visualize the backbone of interactions, we lastly calculated the link overlap of every network given its full set of optimal alignments (Fig. 4*C*). Here, the weight of a link between two species is given by the number of times that link is also shared by those species' pairings across all webs. This allows us to identify sets of links that are consistently aligned across networks—much like we previously identified best aligned species—and to reveal what the backbone of interactions looks like. For a given size k, we identified every network's backbone of interactions made up of the k most-overlapped links. Here, we explored backbones in the range $6 \leqslant k \leqslant 31$, where 6 corresponds to the network with fewest links in our database and the upper bound 31 ensures that we maintained 75% of the networks in the analyses that follow. This analysis of the backbones' overall structure revealed the most common patterns of interaction forming the backbones (Supplementary Fig. 24). Notably, when examining the 118 food webs for which we had interaction strength data (Jacquet et al., 2016), we also observed that the backbones tend to be made up of the strongest links of the community (Supplementary Fig. 25).

To compare the backbones found across food webs, we also aligned them—using the same method as for the full food webs—and generated the corresponding dissimilarity matrix E_k for every backbone size k, where $e_{ij|k}$ is the optimal alignment cost between the k-link

backbones from any network i and j (Eq. 6). Using clustering techniques, we then analyzed these dissimilarity matrices and identified the number of distinct qualitative structures necessary to explain the obtained backbones of interactions (Fig. 7; Methods). Regardless of the size k of the backbones, we found that we could consistently identify two clusters that characterize the observed backbones (Fig. 7). To find the representative structure for each of these clusters, we identified their medoids and generated the respective overlapping structures characterizing each cluster (Fig. 7 and Supplementary Fig. 26). These structures were consistent with the results found using an alternative measure of network similarity, which does not require aligning the backbones ("Links removals: alternative measure for identifying the backbones of interactions" section of Appendix C; Supplementary Fig. 27). Noticeably, we found the differences between the backbones to be less evident when $k > 15$, which could represent a size or detectability limit for the identification of the backbones in our dataset ("Many-link backbones of interactions" section of Appendix C; Supplementary Fig. 28). Finally, we followed the same approach to identify the representative structures for each of the 5 different ecosystem types. Despite showing some expectable variability, the backbones found independently for every ecosystem type largely agree with the backbones found for the entire dataset ("Backbones of interactions for each ecosystem" section of Appendix C; Supplementary Fig. 29).

2.4 *Discussion*

In this paper, we have developed a new approach to align ecological networks as an attempt to shed light on the way those communities resemble each other while avoiding the loss of information associated with comparing derived measures of network structure. Although it has previously been argued that food webs from different environments share a common set of macroscopic properties (Riede et al., 2010; Eklöf et al., 2013), there is also strong evidence suggesting that food-web structure may differ between ecosystems in characteristic ways (Shurin et al., 2006; Vermaat et al., 2009). One way to reconcile these two findings would be if ecological networks presented a backbone of interactions that was shared across environments. That is, ecological networks could all tend to include a set of ecologically equivalent species that almost always interact in a similar fashion while also showing significant differences in the way the remaining species are attached to the periphery.

To test for the existence of this backbone of interactions, we first focused on detecting actual differences on the alignments between networks from different ecosystems. We observed consistent differences in the structure of food webs across ecosystem types. These

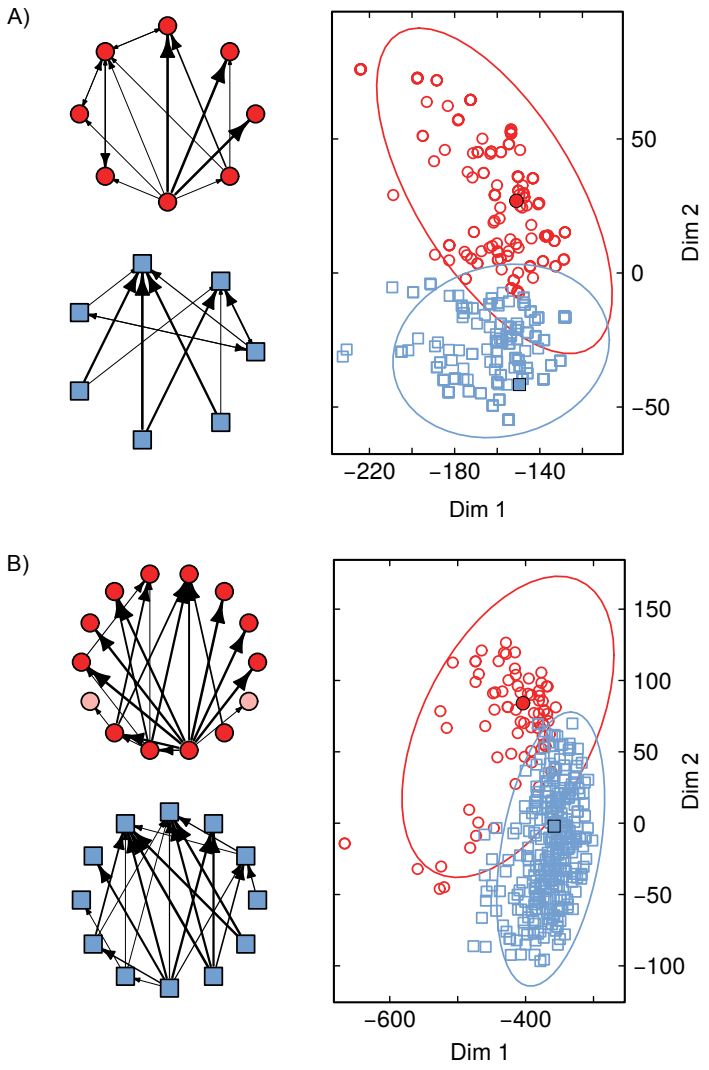


Figure 7: Visualization of the backbones of interactions found across all food webs. (A) Analysis for the 6-link backbones of interactions. On the right, we show a representation of the clustering analysis for the dissimilarity matrix E_6, where every point represents the backbone from a different network. The red and blue network structures depicted on the left characterize the distinct backbones identified within each of the two clusters. They are found by selecting the medoids of the clusters (indicated by the black circle and square) and overlapping them with all the within-cluster backbones, following the example shown in Fig. 4C. In these red and blue structures, the weight of the links is proportional to the likelihood l of finding them in the backbones. Note that links that were not significantly represented in the backbones ($l < 0.01$) are not shown. (B) We show the same analysis but for the 15-link backbones of interactions. The light-red nodes in the top structure indicate nodes that significantly appears in the backbones but not in the medoid.

differences were particularly strong for freshwater stream food webs, which could well indicate different stability mechanisms associated with high seasonal variability (Winterbourn, 1997). It is worth noting, however, that we found measurable structural dissimilarities between almost every pair of ecosystem types. Although some of these dissimilarities may potentially be explained away by differences in the sampling methods used to collect the empirical data in different environments (Martinez, 1991), the strong consistency of our results for such a diverse dataset suggests a fundamental heterogeneity across ecosystems that has rarely been identified previously (Vermaat et al., 2009; Stouffer et al., 2011).

Despite finding consistent differences across ecosystem types, we found that within nearly every network comparison there is a set of species in both food webs that present a better alignment than the rest. Those species are also consistently paired across food webs and far more likely to be connected to each other than would be expected at random. These three results combined hint at the idea that there is indeed a backbone of interactions underpinning all food webs. When examining what this backbone actually looks like, we identified the two most-widespread candidates across all networks. Broadly speaking, the two backbones could be described as follows: a structure with high centrality, where few species in the center that are consumed by many satellite species; and, a "bipartite" structure, where half of the species are consumers of the other half. Despite the observed differences, species forming each backbone do not seem to belong to distinct trophic levels ("Trophic level of the backbones" section of Appendix C). In addition, both backbones were mainly made up of a combination of exploitative competition, generalist predation and simple three-species food chains (Holt and Polis, 1997). As the size of the backbones increases, we also observed an increase in the number of three-species omnivory loops. We advise caution, however, when focusing on the topology found for the backbones. Although their existence is crucial to understanding food-web structure, it doesn't imply that links not found in the backbone are unimportant. Instead, it is best to think that those links are just distributed differently within the networks.

That being said, there are three commonly-studied aspects of food-web structure that could be viewed in a different light given our observations of a consistent backbone. First, even though a backbone could appear to be in contrast with the stabilizing effect associated with compartmentalized food webs (Stouffer and Bascompte, 2011), most of the networks used in this study presented a modular structure ("Compartmentalized structure of food webs" section of Appendix C). This suggests that the observed backbones could exist within mod-

ules, which could explain some of the noise present in our results. Second, the prevalence of omnivory and its role in the stability of food webs has led to equivocal results. While some work has linked the existence of omnivory to lower stability (Tanabe and Namba, 2005; Vandermeer, 2006), there is strong evidence that suggests a positive relationship when trophic interactions are weak (Neutel et al., 2002; Emmerson and Yearsley, 2004). Following this, it is noteworthy that, despite the fact that most of the networks contain three-species omnivory loops, we rarely found this type of interactions within backbones. Regardless of the effect of omnivory interactions on the stability of food webs, this suggests that it is embedded differently across networks. Finally, when considering the networks for which we had interaction strengths, we found that the backbones generally contain the strongest interactions of the community. This may make sense given other correlates of interaction strength. After all, (i) they could otherwise be overlooked in empirical data sets due to sampling errors, and (ii) they might be unable to persist in ecosystems subject to constant environmental change and frequent disturbances.

Among other potential implications of a backbone, we expect that it could be vital to explain and understand food-web dynamics. Similar to the work presented by Murdoch et al. (2002), in which they show that the dynamics of generalist consumers can be approximated using one-species models, the backbone of interactions could also be an internal motor that is driving the dynamics of complex ecological communities. Under this perspective, a backbone of interactions could likewise arise as a potential management tool, whereby the dynamics of entire networks could be optimally regulated by focusing on the species forming the backbone (McDonald-Madden et al., 2016). While, it has been shown that the structure of networks might not necessarily influence their functioning (Allhoff et al., 2015), the backbone could be a driver that ensures at least minimal functioning by staying intact during ongoing species turnover (Allhoff and Drossel, 2016). Along similar lines, these structures could also arise as useful toy models for the study of how ecosystems react to scenarios of current global change (Fussmann et al., 2014). Further inspection of the species attached to the periphery of the backbones, on the other hand, could potentially provide insights into the mechanisms by which food webs from different environments are shaped under different perturbations (Melián and Bascompte, 2002).

This link between structure and dynamics is especially important because measuring and comparing the topology of ecological networks is much easier than elucidating their dynamics, both empirically and synthetically. Although characterizing the properties of ecological networks and identifying their overall differences across environments

have proven to be useful to answer key questions in ecology and evolution (Montoya and Solé, 2003; Allesina and Levine, 2011; Eklöf et al., 2012), aligning ecological networks provides a new level of understanding of "how" exactly ecological networks resemble and differ from each other. Consequently, network alignment presents itself as a powerful and versatile tool for the study of ecological communities. The identification of species that are critically affected by environmental perturbations (Doak et al., 2008), for example, could be used as a strategy for selecting other species from different communities that might be sensitive to similar disturbances. The empirical observation of the dynamics of one ecological network could then be extended to other networks by simply aligning them together, avoiding the use of mathematical models that might oversimplify the dynamics of these ecological systems (Yodzis, 1998; Holland et al., 2002).

Finally, we identify two aspects that stand out as key steps moving forward. First, though computationally intensive, it could be worth testing the existence of a backbone in randomized, as opposed to empirical, communities. These test could reveal the conditions under which different backbone structures emerge (Fortuna et al., 2010). While network properties might significantly change following certain reshuffling processes, backbones could be found to instead persist; it would then be the periphery attached to the backbone that is absorbing the effects of the randomizations (Lu et al., 2016). Second and perhaps more important, further exploration of the species that make up the backbone of interactions should provide a very interesting perspective. If there are indeed intrinsic properties such as traits or shared evolutionary history that are common across the species in the backbone, this could shed light on fundamental aspects of community assembly (Emerson and Gillespie, 2008). Importantly, this might not only untangle the eco-evolutionary mechanisms explaining the formation of such a backbone but may also allow us to understand the role of the backbone as a driver of species' coexistence and diversification (HilleRisLambers et al., 2012).

2.5 *Materials and Methods*

2.5.1 *Empirical data*

We combined the data from multiple previous studies to build a large dataset of networks sampled from different environments and capture as much empirical variation as possible (Cirtwill et al., 2015; Jacquet et al., 2016). Because they are incompatible with our methodology, we excluded any bipartite networks; we also limited ourselves to communities ranging in size from 5 to 133 species due to computational difficulties and greater degeneracies in larger networks. In

total, we used 411 food webs from 34 estuaries, 87 lakes, 148 marine ecosystems, 88 streams and 54 terrestrial ecosystems.

2.5.2 *Species role similarity*

To measure the roles of different species, we used the definition based on the idea of network motifs (Stouffer et al., 2012). Network motifs represent the distinct n-species subnetworks describing all unique patterns of interactions between n species. It has been shown that one can characterize the role of any given species a based on the number of times c_{ai}^{n} that it occupies each distinct position i of the n-species network motifs (Stouffer et al., 2012) ("Alignment algorithm" section of Appendix C). This definition allows a convenient way to compare the topological roles of different species. In particular, given any two species a and b with motif-role profiles $\vec{c}_a$ and $\vec{c}_b$, we used the Pearson's correlation coefficient to define a "measure" of similarity between them as

$$\rho(a,b) = \frac{\operatorname{cov}(\vec{c}_a, \vec{c}_b)}{\sigma_{\vec{c}_a}\sigma_{\vec{c}_b}}, \tag{4}$$

where $\operatorname{cov}(\vec{c}_a, \vec{c}_b)$ is the covariance between roles and $\sigma_{\vec{c}_a}$ and $\sigma_{\vec{c}_b}$ are the standard deviations of $\vec{c}_a$ and $\vec{c}_b$, respectively. This measure of similarity is equal to 1 if a and b play equivalent roles, 0 when there is no correlation between them, and -1 if they play opposite roles.

2.5.3 *Identifying optimal alignments*

We define an alignment between two food webs A and B as a set of one-to-one species pairings $\lambda = \{(a,b)\}$. We allow λ to contain three different types of elements: a unique pairing (a,b) between two species $a \in A$ and $b \in B$; an element $(a,\emptyset)$ representing an unpaired species $a \in A$; and an element $(\emptyset,b)$ representing an unpaired species $b \in B$. Such unpaired species necessarily arise, for example, if the two networks are of different sizes; in addition, species in A and species in B need not resemble each other and hence alignments may not be optimal if dissimilar species are paired together.

Following this definition, the cost function associated with any given alignment can be characterized in multiple ways. One possibility would be to simply consider the sum of every individual species-species pairings as in

$$e_{AB}(\lambda) = \sum_{(a,b)\in\lambda} \big(1 - \rho(a,b)\big). \tag{5}$$

where $\rho(a,b)$ is the measure of role similarity defined above, and for which we assign a penalty of ε for species that remain unpaired

(i.e. $\rho(a, \emptyset) = \rho(\emptyset, b) = \varepsilon$). Minimizing this cost function by changing the alignment λ should directly result in matching species that play similar roles in their respective communities (Supplementary Fig. 30).

Unfortunately, this strategy for optimizing alignments guarantees that similar species from different food webs are matched based on their own structural roles but does not guarantee that their neighbors are optimally matched, or even that their overall networks are aligned ("Alignment algorithm" section of Appendix C). To overcome this drawback, we instead use another cost function to pair up species based on the structural-role similarity of their neighbors. That is, two species from different food webs will only be perfectly matched if their neighbors are also matched with equivalent roles (Supplementary Fig. 30 and Supplementary Fig. 31). Therefore, the contribution of two paired species to the overall cost function will be the sum across their neighbors' pairings. With this in mind, we define an improved cost function as follows:

$$e_{AB}(\lambda) = \sum_{x \in \lambda} \left(\sum_{(\alpha,\beta) \in \lambda_x} \big(1 - \rho(\alpha, \beta)\big) + \xi_x \right), \tag{6}$$

where, given the pairing $x = (a, b)$ between two species $a \in A$ and $b \in B$, we define the subset $\lambda_x = \lambda_{(a,b)}$ of λ as the set of all the one-to-one pairing (α, β) containing both a neighbor α of a and a neighbor β of b. Following this, ξ_x represents the penalty associated with the unpaired neighbors of every pairing $x = (a, b)$, which accounts for both the number of neighbors of a that are not paired with a neighbor of b and the number of neighbors of b that are not paired with a neighbor of a ("Alignment algorithm" section of Appendix C).

2.5.4 *Alignment quality*

In order for the alignments to be comparable across our dataset, we also need a network-size-independent measure of how good those alignments are. This is because the alignment cost function defined above is useful for optimizing pairwise network alignments but strongly scales with the size of the networks being aligned. Although neutralizing this size effect is nontrivial, there are multiple ways to appropriately reduce the effect of a size difference between networks ("Alignment quality measures" section of Appendix C). Here, we adopt an approach described as follows. Given the best alignment $\hat{\lambda}$ found between two networks A and B, we calculate the normalized dissimilarity $\hat{e}_{AB}(\hat{\lambda})$ between them rewriting Eq. (5) as follows:

$$\hat{e}_{AB}(\hat{\lambda}) = \frac{1}{N} \sum_{(a,b) \in \lambda} \big(1 - \rho(a, b)\big), \tag{7}$$

where we now set the cost associated with an unpaired species to $\rho(a, \emptyset) = \rho(\emptyset, b) = 1$, and N represents the total number of matches between one species from A and one species from B. We chose a normalized version of Eq. (5) for alignment quality because it is much simpler than the same for Eq. (6). Other alignment quality measures are also considered in the "Alignment quality measures" section of Appendix C, Supplementary Figure 32.

2.5.5 *Quantifying structural differences across ecosystems*

To test for differences across ecosystem types, we analyzed the alignment dissimilarity matrices using a permutational multivariate analysis of variance (Anderson, 2001) (PERMANOVA), which expands beyond the traditional analysis of variance methods (ANOVA) and assesses relative differences between and within treatment groups (e.g. ecosystem types) using a permutation-based significance test.

2.5.6 *Alignment transitivity*

The transitivity between alignments characterizes the cliquishness of all species-species alignments. Suppose that we align a set of food webs $\{A, B, \ldots, Z\}$. Given that species $a \in A$ is aligned with species $b \in B$ and $c \in C$, the alignment transitivity of a is the likelihood of b and c also being aligned.

2.5.7 *Path likelihood*

The path likelihood is a useful measure for testing whether or not a set of species of a network form a connected component. Given a network A comprised of n species, the path probability of a subset comprised of $k < n$ species is defined as the probability that at least one undirected path existed between all pairs of k species.

2.5.8 *Number of distinct backbones*

To find the number of different candidate backbones, we used the *R* package *NbClust*, which determines the number of clusters that characterize a dissimilarity matrix by means of combining 5 different indices and 8 clustering methods (Charrad et al., 2014). Given the number of clusters from each index and method, we used the majority rule to identify the actual number of clusters.

2.5.9 *Code availability*

Code to conduct the network alignment described here can be made available upon request.

2.5.10 *Data availability*

Data to conduct the analyses performed here can be obtained following Cirtwill et al. (2015) and Jacquet et al. (2016), or made available upon request.

BIBLIOGRAPHY

Marcelo A Aizen, Carolina L Morales, and Juan M Morales. Invasive mutualists erode native pollination webs. *PLoS Biology*, 6(2):e31, 2008.

Marcelo A Aizen, Malena Sabatino, and Jason M Tylianakis. Specialization and rarity predict nonrandom loss of interactions from mutualist networks. *Science*, 335(6075):1486–1489, 2012.

Stefano Allesina and Jonathan M Levine. A competitive network theory of species diversity. *Proceedings of the National Academy of Sciences*, 108(14):5638–5642, 2011.

Stefano Allesina and Si Tang. Stability criteria for complex ecosystems. *Nature*, 483(7388):205–208, 2012.

Stefano Allesina, Jacopo Grilli, György Barabás, Si Tang, Johnatan Aljadeff, and Amos Maritan. Predicting the stability of large structured food webs. *Nature Communications*, 6:7842, 2015.

Korinna Theresa Allhoff and Barbara Drossel. Biodiversity and ecosystem functioning in evolving food webs. *Philosophical Transactions of the Royal Society B*, 371(1694):20150281, 2016.

Korinna Theresa Allhoff, Daniel Ritterskamp, Björn C Rall, Barbara Drossel, and Christian Guill. Evolutionary food web model based on body masses gives realistic networks with permanent species turnover. *Scientific Reports*, 5(10955), 2015.

Marti J Anderson. A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26(1):32–46, 2001.

Benjamin Baiser, Nicholas J Gotelli, Hannah L Buckley, Thomas E Miller, and Aaron M Ellison. Geographic variation in network structure of a nearctic aquatic food web. *Global Ecology and Biogeography*, 21(5):579–591, 2012.

Nico Blüthgen. Why network analysis is often disconnected from community ecology: a critique and an ecologist's guide. *Basic and Applied Ecology*, 11(3):185–195, 2010.

Jonathan J Borrelli, Stefano Allesina, Priyanga Amarasekare, Roger Arditi, Ivan Chase, John Damuth, Robert D Holt, Dmitrii O Lo-

gofet, Mark Novak, Rudolf P Rohr, et al. Selection on stability across ecological scales. *Trends in Ecology and Evolution*, 30(7):417–425, 2015.

James H Brown, James F Gillooly, Andrew P Allen, Van M Savage, and Geoffrey B West. Toward a metabolic theory of ecology. *Ecology*, 85(7):1771–1789, 2004.

Malika Charrad, Nadia Ghazzali, Veronique Boiteau, and Azam Niknafs. Package 'nbclust'. *Journal of Statistical Software*, 61:1–36, 2014.

Alyssa R Cirtwill and Daniel B Stouffer. Knowledge of predator–prey interactions improves predictions of immigration and extinction in island biogeography. *Global Ecology and Biogeography*, 25(7):900–911, 2015.

Alyssa R Cirtwill, Daniel B Stouffer, and Tamara N Romanuk. Latitudinal gradients in biotic niche breadth vary across ecosystem types. *Proceedings of the Royal Society B: Biological Sciences*, 282(1819): 20151589, 2015.

David J Currie, Gary G Mittelbach, Howard V Cornell, Richard Field, Jean-Francois Guégan, Bradford A Hawkins, Dawn M Kaufman, Jeremy T Kerr, Thierry Oberdorff, Eileen O'Brien, et al. Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecology Letters*, 7(12):1121–1134, 2004.

Daniel F Doak, James A Estes, Benjamin S Halpern, Ute Jacob, David R Lindberg, James Lovvorn, Daniel H Monson, M Timothy Tinker, Terrie M Williams, J Timothy Wootton, et al. Understanding and predicting ecological dynamics: are major surprises inevitable. *Ecology*, 89(4):952–961, 2008.

Jennifer A Dunne, Richard J Williams, and Neo D Martinez. Food-web structure and network theory: the role of connectance and size. *Proceedings of the National Academy of Sciences*, 99(20):12917–12922, 2002a.

Jennifer A Dunne, Richard J Williams, and Neo D Martinez. Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecology Letters*, 5(4):558–567, 2002b.

Anna Eklöf, Matthew R Helmus, M Moore, and Stefano Allesina. Relevance of evolutionary history for food web structure. *Proceedings of the Royal Society B: Biological Sciences*, 279(1733):1588–1596, 2012.

Anna Eklöf, Ute Jacob, Jason Kopp, Jordi Bosch, Rocío Castro-Urgal, Natacha P Chacoff, Bo Dalsgaard, Claudio Sassi, Mauro Galetti, Paulo R Guimarães, et al. The dimensionality of ecological networks. *Ecology Letters*, 16(5):577–583, 2013.

Brent C Emerson and Rosemary G Gillespie. Phylogenetic analysis of community assembly and structure over space and time. *Trends in Ecology and Evolution*, 23(11):619–630, 2008.

Mark Emmerson and Jon M Yearsley. Weak interactions, omnivory and emergent food-web properties. *Proceedings of the Royal Society B: Biological Sciences*, 271(1537):397–405, 2004.

Jason Flannick, Antal Novak, Balaji S Srinivasan, Harley H McAdams, and Serafim Batzoglou. Graemlin: general and robust alignment of multiple large interaction networks. *Genome Research*, 16(9):1169–1181, 2006.

Miguel A Fortuna, Daniel B Stouffer, Jens M Olesen, Pedro Jordano, David Mouillot, Boris R Krasnov, Robert Poulin, and Jordi Bascompte. Nestedness versus modularity in ecological networks: two sides of the same coin? *Journal of Animal Ecology*, 79(4):811–817, 2010.

Katarina E Fussmann, Florian Schwarzmüller, Ulrich Brose, Alexandre Jousset, and Bjoern C Rall. Ecological stability in response to warming. *Nature Climate Change*, 4(3):206–210, 2014.

Lloyd Goldwasser and Jonathan Roughgarden. Sampling effects and the estimation of food-web properties. *Ecology*, 78(1):41–54, 1997.

Dominique Gravel, Elsa Canard, Frédéric Guichard, and Nicolas Mouquet. Persistence increases with diversity and connectance in trophic metacommunities. *PLoS ONE*, 6(5):e19374, 2011a.

Dominique Gravel, Francois Massol, Elsa Canard, David Mouillot, and Nicolas Mouquet. Trophic theory of island biogeography. *Ecology Letters*, 14(10):1010–1016, 2011b.

Stephen J Hall and Dave G Raffaelli. *Food webs: Theory and Reality*, volume 24. Academic Press, 1993.

Ian A Hatton, Kevin S McCann, John M Fryxell, T Jonathan Davies, Matteo Smerlak, Anthony RE Sinclair, and Michel Loreau. The predator-prey power law: Biomass scaling across terrestrial and aquatic biomes. *Science*, 349(6252):aac6284, 2015.

Janneke HilleRisLambers, Peter B Adler, William S Harpole, Jonathan M Levine, and Margaret M Mayfield. Rethinking community assembly through the lens of coexistence theory. *Annual Review of Ecology, Evolution, and Systematics*, 43:227–248, 2012.

Julian N Holland, Donald L DeAngelis, and Judith L Bronstein. Population dynamics and mutualism: functional responses of benefits and costs. *The American Naturalist*, 159(3):231–244, 2002.

Robert D Holt. *The theory of island biogeography revisited*, chapter Toward a trophic island biogeography, pages 143–185. Princeton University Press, Princeton, 2010.

Robert D Holt and Gary A Polis. A theoretical framework for intraguild predation. *The American Naturalist*, 149(4):745–764, 1997.

Claire Jacquet, Charlotte Moritz, Lyne Morissette, Pierre Legagneux, François Massol, Philippe Archambault, and Dominique Gravel. No complexity-stability relationship in empirical ecosystems. *Nature Communications*, 7:12573, 2016.

Oleksii Kuchaiev, Tijana Milenković, Vesna Memišević, Wayne Hayes, and Nataša Pržulj. Topological network alignment uncovers biological function and phylogeny. *Journal of the Royal Society Interface*, 7(50):1341–1354, 2010.

John H Lawton. *Ecological Concepts*, chapter Food webs. Blackwell Scientific, 1989.

Xueke Lu, Clare Gray, Lee E Brown, Mark E Ledger, Alexander M Milner, Raúl J Mondragón, Guy Woodward, and Athen Ma. Drought rewires the cores of food webs. *Nature Climate Change*, 6:875–878, 2016.

Neo D Martinez. Artifacts or attributes? effects of resolution on the little rock lake food web. *Ecological Monographs*, 61(4):367–392, 1991.

Eve McDonald-Madden, Régis Sabbadin, Edward T Game, Peter WJ Baxter, Iadine Chadès, and Hugh Possingham. Using food-web theory to conserve ecosystems. *Nature Communications*, 7:10245, 2016.

Carlos J Melián and Jordi Bascompte. Complex networks: two ways to be robust? *Ecology Letters*, 5(6):705–708, 2002.

Jose M Montoya and Ricard V Solé. Topological properties of food webs: from real data to community assembly models. *Oikos*, 102 (3):614–622, 2003.

Rebecca J Morris. Anthropogenic impacts on tropical forest biodiversity: a network structure and ecosystem functioning perspective. *Philosophical Transactions of the Royal Society B*, 365(1558):3709–3718, 2010.

William W Murdoch, Bruce E Kendall, Roger M Nisbet, Cheryl J Briggs, Edward McCauley, and Robin Bolser. Single-species models for many-species food webs. *Nature*, 417(6888):541–543, 2002.

Anje-Margriet Neutel, Johan AP Heesterbeek, and Peter C de Ruiter. Stability in real food webs: weak links in long loops. *Science*, 296 (5570):1120–1123, 2002.

Rob Patro and Carl Kingsford. Global network alignment using multiscale spectral signatures. *Bioinformatics*, 28(23):3105–3114, 2012.

Owen L Petchey, Ulrich Brose, and Björn C Rall. Predicting the effects of temperature on food web connectance. *Philosophical Transactions of the Royal Society B*, 365(1549):2081–2091, 2010.

Denise A Piechnik, Sharon P Lawler, and Neo D Martinez. Food-web assembly during a classic biogeographic study: species'"trophic breadth" corresponds to colonization order. *Oikos*, 117(5):665–674, 2008.

Timothée Poisot, Elsa Canard, David Mouillot, Nicolas Mouquet, and Dominique Gravel. The dissimilarity of species interaction networks. *Ecology Letters*, 15(12):1353–1361, 2012.

Timothée Poisot, Daniel B Stouffer, and Dominique Gravel. Beyond species: why ecological interaction networks vary through space and time. *Oikos*, 124(3):243–251, 2015.

Jens O Riede, Björn C Rall, Carolin Banasek-Richter, Sergio A Navarrete, Evie A Wieters, Mark C Emmerson, Ute Jacob, and Ulrich Brose. Scaling of food-web properties with diversity and complexity across ecosystems. *Advances in Ecological Research*, 42:139–170, 2010.

Serguei Saavedra, Daniel B Stouffer, Brian Uzzi, and Jordi Bascompte. Strong contributors to network persistence are the most vulnerable to extinction. *Nature*, 478(7368):233–235, 2011.

Jonathan B Shurin, Daniel S Gruner, and Helmut Hillebrand. All wet or dried up? real differences between aquatic and terrestrial food webs. *Proceedings of the Royal Society B: Biological Sciences*, 273(1582): 1–9, 2006.

Rohit Singh, Jinbo Xu, and Bonnie Berger. Pairwise global alignment of protein interaction networks by matching neighborhood topology. In *Annual International Conference on Research in Computational Molecular Biology*, pages 16–31. Springer, 2007.

Daniel B Stouffer and Jordi Bascompte. Understanding food-web persistence from local to global scales. *Ecology Letters*, 13(2):154–161, 2010.

Daniel B Stouffer and Jordi Bascompte. Compartmentalization increases food-web persistence. *Proceedings of the National Academy of Sciences*, 108(9):3648–3652, 2011.

Daniel B Stouffer, Juan Camacho, Wenxin Jiang, and Luís A Nunes Amaral. Evidence for the existence of a robust pattern of prey selection in food webs. *Proceedings of the Royal Society B: Biological Sciences*, 274(1621):1931–1940, 2007.

Daniel B Stouffer, Enrico L Rezende, and Luís A Nunes Amaral. The role of body mass in diet contiguity and food-web structure. *Journal of Animal Ecology*, 80(3):632–639, 2011.

Daniel B Stouffer, Marta Sales-Pardo, M Irmak Sirer, and Jordi Bascompte. Evolutionary conservation of species' roles in food webs. *Science*, 335(6075):1489–1492, 2012.

Kumi Tanabe and Toshiyuki Namba. Omnivory creates chaos in simple food web models. *Ecology*, 86(12):3411–3414, 2005.

Elisa Thébault and Colin Fontaine. Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science*, 329(5993):853–856, 2010.

Ross M Thompson and Colin R Townsend. Energy availability, spatial heterogeneity and ecosystem size predict food-web structure in streams. *Oikos*, 108(1):137–148, 2005.

John Vandermeer. Omnivory and the stability of food webs. *Journal of Theoretical Biology*, 238(3):497–504, 2006.

Jan E Vermaat, Jennifer A Dunne, and Alison J Gilbert. Major dimensions in food-web structure properties. *Ecology*, 90(1):278–282, 2009.

Richard J Williams. Simple maxent models explain food web degree distributions. *Theoretical Ecology*, 3(1):45–52, 2010.

Richard J Williams and Neo D Martinez. Simple rules yield complex food webs. *Nature*, 404(6774):180–183, 2000.

Richard J Williams, Eric L Berlow, Jennifer A Dunne, Albert-László Barabási, and Neo D Martinez. Two degrees of separation in complex food webs. *Proceedings of the National Academy of Sciences*, 99 (20):12913–12916, 2002.

Michael J Winterbourn. *Evolutionary Ecology of Freshwater Animals*, chapter New Zealand mountain stream communities: Stable yet disturbed?, pages 31–54. Springer, 1997.

Peter Yodzis. Local trophodynamics and the interaction of marine mammals and fisheries in the benguela ecosystem. *Journal of Animal Ecology*, 67(4):635–658, 1998.

3

PYMFINDER: A TOOL FOR THE MOTIF ANALYSIS OF BINARY AND QUANTITATIVE COMPLEX NETWORKS

Bernat Bramon Mora[1], Alyssa R. Cirtwill[2], Daniel B. Stouffer[1]

1 Centre for Integrative Ecology, School of Biological Sciences, University of Canterbury, Christchurch, New Zealand

2 Department of Physics, Chemistry, and Biology, Linköping University, Linköping, Sweden.

3.1 *Abstract*

We developed *pymfinder*, a new software to analyze multiple aspects of the so-called network motifs—distinct n-node patterns of interaction—for any directed, undirected, unipartite or bipartite network. Unlike existing software for the study of network motifs, *pymfinder* allows the computation of node- and link-specific motif profiles as well as the analysis of weighted motifs. Beyond the overall characterization of networks, the tools presented in this work therefore allow for the comparison of the "roles" of either nodes or links of a network. Examples include the study of the roles of different species and/or their trophic/mutualistic interactions in ecological networks or the roles of specific proteins and/or their activation/inhibition relationships in protein-protein interaction networks. Here, we show how to apply the main tools from *pymfinder* using a predator-prey interaction network from a marine food web. *pymfinder* is open source software that can be freely and anonymously downloaded from `https://github.com/stoufferlab/pymfinder`, distributed under the MIT License (2018).

3.2 *Introduction*

The use of network theory has proven insightful in multiple fields, from the study of the spread of disease epidemics (Newman, 2002) to the characterization of neuronal networks (Sporns, 2002). In ecology, this approach has been crucial to understanding the ways different species interact with each other, and the network perspective has justly become a central topic in community ecology (Bascompte and Jordano, 2013). Over recent years, multiple methods for studying

the topology of ecological networks have been successfully developed. Examples include models to generate realistic ecological communities (Williams and Martinez, 2000) or tools for studying different network metrics such as compartmentalization (Guimera et al., 2007), nestedness (Bascompte et al., 2003) or intervality (Stouffer et al., 2006). Following these advances, one of the most versatile ways to understand the structure of complex ecological networks is via the so-called network motifs—i.e. the analysis of small subgraphs representing the distinct patterns of interaction involving any set of n species. These subgraphs have been referred to as the 'building blocks' of complex networks (Milo et al., 2002).

The study of network motifs has been applied to multiple ecological systems over the recent years, including those composed of trophic (Stouffer et al., 2007) and mutualistic interactions (Dormann et al., 2009). Non-ecological examples include in protein-protein interaction networks (Yeger-Lotem et al., 2004) and transcriptional regulation networks (Shen-Orr et al., 2002). There are typically two main approaches that are taken involving network motifs. First, counting the number of appearances of any given n-node pattern of interactions provides an overall perspective of the structure of a network. This has been done in different ecological studies, including the characterization of food webs (Klaise and Johnson, 2017), plant-pollination (Rodríguez-Rodríguez et al., 2017) and host-parasitoid networks (Baker et al., 2015). Second, other ecological studies have focused on the role of different species (Stouffer et al., 2012) and interactions (Cirtwill and Stouffer, 2015), defining their position within the network based on which network motifs they form a part of. Following this work on network motifs, multiple tools for the counting of network motifs have been developed over the last decades (Kashtan et al., 2004; Wernicke and Rasche, 2006; Csardi and Nepusz, 2006). Most of the methodological work has focused on providing tools to efficiently quantify the overall structure of directed and undirected unipartite networks—i.e. graphs consisting of one set of interacting nodes. Unfortunately, to our knowledge, we are still lacking general-purpose software to also analyze bipartite networks—i.e. graphs consisting of two interacting sets of non-overlapping nodes—as well as to quantify the node- and link-specific motif profiles in both unipartite and bipartite networks. In addition, there is no tool to date that allows the user to include information regarding the interaction strengths of a network within the analysis of motifs. In response, we present *pymfinder*, software for motif analysis of network structure plus of the nodes and links of any type of network—i.e. directed/undirected, bipartite/unipartite, and weighted/binary networks.

pymfinder is an open-source and versatile tool for the study of network motifs and the result of long-standing research involving the study of ecological networks. For example, *pymfinder* was used to shed light on the ecological mechanisms underlying food-web structure (Stouffer et al., 2007), which, together with Bascompte 2005 (Bascompte and Melián, 2005) and Camacho et. al. (Camacho et al., 2007), was one of the first studies to put network motifs into a purely ecological context. Building on these foundational studies, network motifs and *pymfinder* were shown to provide a useful way to characterize species' roles, showing them to be evolutionary conserved across communities (Stouffer et al., 2012). Similarly, the roles of links involving parasite species were characterized through the study of network motifs, generating an understanding of how different types of feeding links are distributed within a food web (Cirtwill and Stouffer, 2015). The same software has also been used to study bipartite networks. For instance, a study on host-parasitoid networks showed how species' roles seem to be conserved over spatial scales as well as consistent over time (Baker et al., 2015). Perhaps more importantly, the software presented here has also been a central piece of very recent research. For example, the tools in *pymfinder* were used to relate species' roles to multiple ecological traits in five marine food webs, showing that feeding environment is particularly strongly related to such roles (Cirtwill and Eklöf, 2018). Likewise, the variability of species' roles in plant-pollinator communities in the Arctic has recently shown to be related to the variability in community composition (Cirtwill et al., 2018). Finally, the description of species' roles has also been key to comparing entire networks by means of aligning species to each other, resulting in the identification of common backbones shared across food webs form different ecosystems (Bramon Mora et al., 2018). Overall, the tools included in *pymfinder* are and have been instrumental to the development of a diverse set of projects over the years, and we believe that they have the potential to be valuable for many others. This article describes the main structure of *pymfinder* and showcases some of its principal applications using two different biological datasets as the backdrop.

3.3 *Design and implementation*

pymfinder is a Python library that combines Python methods for network-motif analysis. Some of the engine underneath is a modified version of *mfinder*—a software tool for network-motif detection developed by Kashtan et. al. (Milo et al., 2002; Kashtan et al., 2004). Originally, *mfinder* was written in C and made available solely as an executable, and we use it within *pymfinder* for its underlying efficiency. The *mfinder* code has been both included and modified here with the explicit consent of Nadav Kashtan, the author of mfinder 1.2.

3.3.1 *General description*

As input, *pymfinder* accepts any type of network. That is, the analyses can be performed for both unipartite and bipartite networks. The format in which the networks are passed to the different functions of the package is either as text files, Python arrays or pymfinder-objects. Text files must describe the set of links comprising the networks, where each link appears as a separate line in the files. For example, a given line "A B w" would describe a single link A $\rightarrow$ B between nodes A and B, where w represents the strength or weight associated to such link (see Appendix). Similarly, Python arrays need to represent the list of interactions forming the networks. Notice that the direction of the links is important. Therefore, in bipartite networks, nodes of each group need to consistently be placed on the same side of the interactions—e.g. in a plant-pollinator networks the direction of the interactions in the input must all go from a plant to a pollinator (or vice-versa). Importantly, undirected networks can also be analyzed by *pymfinder*; however, any links between two nodes A and B in such networks need to be characterized by the two parallel links A B and B A. The output of *pymfinder*, is a high-level data type ('class') that contains different descriptors of the motif composition of the network under study (see Appendix).

3.3.2 *Structure of the package*

At their core, all of the analyses performed by *pymfinder* are based around the identification of all the different n-node patterns of interaction found within a given network. To do this, *pymfinder* will always start by enumerating the unique motifs/subgraphs that make up the overall structure of the network under study. This analysis can be performed for multiple motif sizes. This is especially important for bipartite networks, where three-node motifs are minimally informative and one needs to explore bigger motifs (Baker et al., 2015). Notice, however, that increasing the number of nodes can be computationally challenging for unipartite networks since the number of unique motifs quickly increases with their size—i.e. there are 13 unique three-node motifs, 199 unique four-node motifs and 9364 unique five-node motifs.

For the sake of simplicity, we will focus most of the description of the methods presented here on the analysis of three-node network motifs. For any given network, this analysis is a three-step process. First, *pymfinder* loops through all the rows i of the adjacency matrix A associated with the network. For each non-zero element a_{ij} found in row A_i, it then searches for any connected element $a_{jk} = 1$, $a_{kj} = 1$, $a_{ik} = 1$, and/or $a_{ki} = 1$, revealing the existence of any motif comprised of the nodes i, j, and k. If i, j and k define a motif and this

motif has not already been identified, the corresponding motif and the position of each node within the motif is recorded.

Based on this initial motif enumeration, *pymfinder* can perform three subsequent analyses: (i) the analysis of the overall network structure, (ii) the nodes and links' participation in the different motifs, and (iii) the nodes and links' role in each of the motifs.

3.3.2.1 *Motif structure*

The most basic application of *pymfinder* is the analysis of the overall motif structure of a given network. In particular, such analysis generates a description of the distribution of distinct n-node patterns of interaction found within the network (up to 8-node motifs). The application also includes the possibility of estimating the null motif composition expected for such network (see Appendix). To generate this null composition, *pymfinder* uses an MCMC algorithm to perform a randomization of the network while preserving the in- and out-degree of the nodes and each node's number of single and double links (Milo et al., 2003; Sanderson et al., 1998). Comparing the observed motif frequency to the random expectation, the application can be used to determine which interaction patterns are over- or under-represented relative to this null model (Stouffer et al., 2007). To do so, *pymfinder* calculates the mean and standard deviation of the null expectation as well as the z-scores for its comparison with the actual observations.

An additional feature of *pymfinder* is the possibility of incorporating information regarding the link strength into the analysis of the motif structure. This is notable in particular since there is no software available to explore the way the interaction strengths are distributed within networks across motifs. To do so, *pymfinder* will account for each motif within a given weighted network as a function of the strength of the links forming them (Fig. 8). Note that the algorithm allows the user to choose how the weight of a motif is defined. Specifically, given a motif formed by the set of links with strengths $\{l\} = \{l_1, l_2, ..., l_L\}$, *pymfinder* will calculate the weight of such motif as $f(\{l\})$, where f is the function defined by the user. By default, *pymfinder* uses the arithmetic mean as the function f. Similar to unweighted networks, analysis of the motif structure of a weighted network returns the average and standard deviation of the weight of each motif, as well as the median and the first and third quartiles.

3.3.2.2 *Motif participation*

The study of network motifs can also be used as a way to classify nodes based on which patterns of interactions they are part of. For any given network, this application determines the frequency of ap-

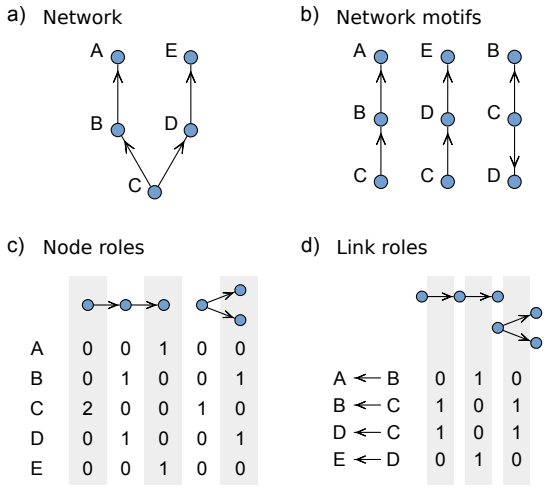


Figure 8: Main components of network-motif analysis. (a) A simple network that could represent a simple ecological community—where nodes would characterize species and the arrows would indicate the interactions between them—-or a protein-protein interaction network—where nodes would represent different proteins and the arrows indicate either activation or inhibition. (b) All three-node motifs found in the network from (a); from this classification, we can compute the overall network structure and the number appearances of every node in each motif. (c) The characterization of every node's motif-role profile. This characterization is based on the number of appearances of every node in each of the unique motif node-positions. (d) The characterization of every link's motif-role profile, which is based on the number of appearances of every link in each of the unique motif link-positions. Notice that we excluded any motif or role that was not represented in the network.

pearance of every node across each of the different motifs (Fig. 8b), defining their participation across these distinct patterns of interactions. This a useful perspective for motif analysis because it provides a node-based description of the networks that can be used to understand the nature of specific nodes (e.g. different species in ecological networks or different proteins in protein-protein interaction networks) as well as decomposes the overall structure of the network at a finer resolution (Bascompte and Melián, 2005). Similarly, the same analysis can also be performed for the links forming the network. That is, *pymfinder* can quantify the frequency with which every link forms part of each distinct motif. As for the analysis of the overall structure of the networks, the motif participation of both nodes and links can also be calculated for any given motif size up to 8 nodes for weighted and unweighted networks. Again, *pymfinder* will account for each motif within a given weighted network as a function of the strength of the links forming it (Fig. 8), and the algorithm allows the

user to choose this function just as described above for motif structure.

3.3.2.3 *Motif-role profiles*

Within any given motif, nodes can play multiple roles. For example, in the two-node motif $A \rightarrow B$, there are two distinct positions A and B, which define two different roles—e.g. a predator and a prey in a food web. In contrast, for the two-node motif $A \leftrightarrow B$, A and B occupy indistinguishable positions; therefore, there is a single distinct role. The same idea can be extended to all n-node motifs. For example, there are 30 distinct node positions and 24 distinct link positions across the 13 unique three-nodes motifs. These distinct positions within the different motifs are important because the number of times that a node appears in each of them can be used as a way to define its structural role in a community (Stouffer et al., 2012). That is, we can characterize a node's structural role based on the number of times that it occupies each distinct position of the n-node motifs. *pymfinder* provides a way to determine such n-node motif-role profiles for both the nodes (Fig. 8c) and the links (Fig. 8d) of a given network. Notice, however, that this function can only be run for two- and three-nodes motifs in unipartite networks, and two- to six-nodes motifs in bipartite networks.

The analysis of node and link motif-role profiles can also incorporate information regarding the strengths of interactions between nodes. As before, consider a motif m formed by the set of nodes $\{i\}$ and the set of links with strengths $\{l\}$. For any node j in $\{i\}$, *pymfinder* calculates the contribution c_{jm} of motif m to any of the positions of j's motif-role profile as:

$$c_{jm} = \frac{f(\{l_j\})}{\sum_i f(\{l_i\})} f(\{l\}) \tag{8}$$

where $\{l_i\}$ is the set of strengths of all links in m involving node i, and f is a function defined by the user. By default, *pymfinder* again uses the arithmetic mean as f for weighted motif-role profiles. Notice that the contribution $c_{im} = 1$ when ignoring the weights, or f is the arithmetic mean and all weights are equal to the motif size. When analyzing the motif-role profile of a link k forming such motif, the contribution c_{km} is assumed to be exactly equal to its link strength l_k.

3.3.3 *Basic tests*

To ensure the reliable functioning of *pymfinder*, we included a set of basic tests in the package. All these basic tests are based around the idea of analyzing the structure of artificial networks containing only a single motif of each type for a given motif size—up to five-node

motifs for bipartite networks and three-node motifs for unipartite networks. In addition, those networks are also set up so that any given node or link is only involved in a single motif and role. Using these single-motif networks, we tested the functions of *pymfinder* by ensuring that the analysis of such artificial networks does not result in the misrepresentation of any motif, node, link or role.

3.4 *Results/Discussion*

The tools provided by *pymfinder* can be used in a large variety of systems and do not depend on the nature or providence of the networks. To illustrate the capabilities and potential of the software, we outline the study of a food web from a marine ecosystem as a representative study system (Bascompte et al., 2005). This specific network describes the predator-prey interactions between approximately 250 of the species found across an extensive area of the Caribbean Sea.

We first analyzed the overall three-species motif structure of the network and compared it to the random expectation (Fig. 9b). For this example, we used the z-score values to draw this comparison, which assume normality of the motif distribution. Notice, however, that *pymfinder* also returns the mean number of motif counts in the randomized networks, which allows for other types of statistical analyses. We found that the observed motif distribution is generally significantly different from the random expectation, showing either over- or under-representation relative to the results of the null model used here. This is evidence of a non-random organization of ecological communities (Bascompte and Melián, 2005; Borrelli, 2015), which speaks to the eco-evolutionary mechanisms shaping the ways in which different species interact with each other. We then studied the distribution of link weights across motifs to test whether or not different motifs are generally made of different interaction strengths. For this particular example, we log-transformed the link weights to be approximately normally distributed as well as scaled them so that the weakest and strongest links had a weight of zero and one, respectively. In general, we found that interaction strengths are distributed in a similar manner across the different motifs of the network under study (Fig. 9c). Notice that these results are subject to the logarithmic transformation applied to the weight data, which is generally very skewed (Bascompte et al., 2005).

Following the analysis of the overall motif structure, we examined the motif participation of the different nodes and links that make up this food web. We found that some nodes (e.g., sea cucumbers and algae) share almost identical motif-participation profiles while others (e.g., filefish and sea cucumbers) have very distinct profiles (Fig. 10a). This

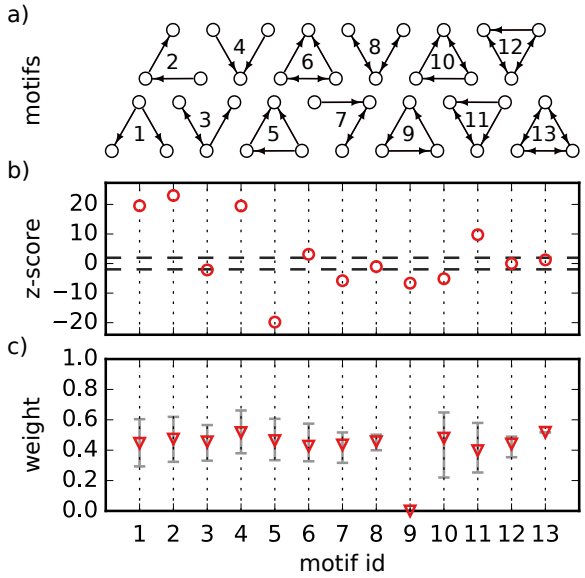


Figure 9: Analysis of the overall motif structure of the marine food web under study. The first panel (a) shows all the possible three-species motif structures. In this case, any arrow indicates the direction of energy flow from a prey to its predator. The second panel (b) presents the z-scores obtained from comparison between the empirical motif frequency and the random expectation. The dotted lines indicate the thresholds for significant over- and under-representation ($z = 1.97$ and $z = -1.97$, respectively). The third panel (c) shows the median weight found for each motif. The error bars represent the first and third quartiles. Note that the motif id given on the x-axis corresponds to the indexing in (a), and that the interaction strengths have been transformed to approximately be normally distributed and strictly positive.

shows how motifs can be a valuable and insightful way to classify and compare the species across communities. Perhaps more importantly, we observed how the information regarding the interaction strengths forming the motifs changed those motif-participation profiles (Fig. 10b). Therefore, adding interaction strengths allowed us to distinguish between the roles of species with similar unweighted profiles. This is important because it suggests that, from a node-specific perspective, interaction strengths are not equally distributed across motifs. The uneven distribution of interaction strengths has important implications for the relationship between network structure and species–interaction strengths and the stability of food webs(Neutel et al., 2002; Emmerson and Yearsley, 2004). We also looked at the motif-participation profiles of the links (Fig. 10c). We found that those profiles could also be an indicator of the observed differences on the way interaction strengths are distributed across motifs, as suggested by previous work (Cirtwill and Stouffer, 2015).

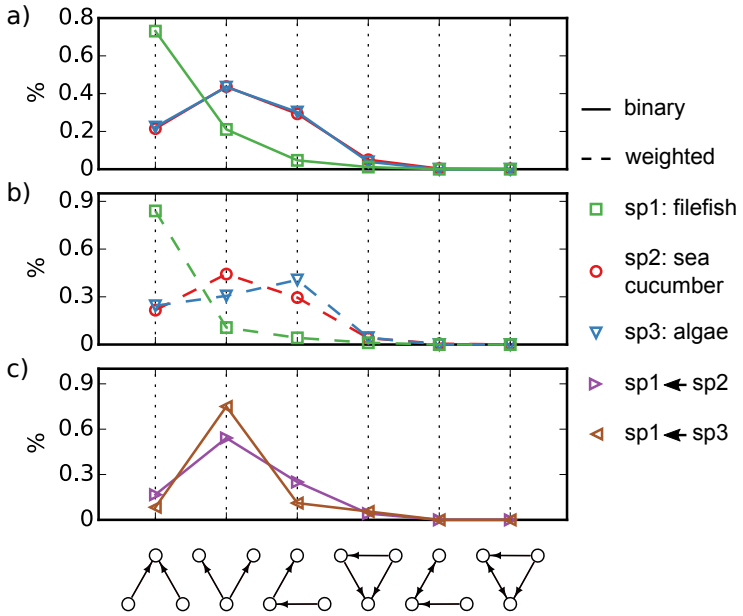


Figure 10: Analysis of the species' motif participation in the marine food web under study. The first panel (a) shows the motif-participation profiles of three representative species from the web; here, every point describes the proportion of times that these species are found in any of the possible motifs. For simplicity, we excluded the seven motifs in which these species never appear. The second panel (b) presents the motif-role profiles for the same three species when adding information regarding the interaction strengths. In this case, every point represents the relative weight associated with the motifs in which each species participates. The third panel (c) shows the motif-participation profiles for the links involving the same three species.

Finally, we studied the motif-role profiles of the species of the marine network. This analysis is similar to the motif participation analysis of nodes and links; however, it provides a finer resolution to the role that different species or links might play in the community. Using the proportion of times that the different species are in each of the 30 unique positions of the three-species network motifs, we performed an analysis of multivariate homogeneity of group dispersions to compare the roles of the species in the network (Anderson, 2006). To do this, we first calculated the euclidean distance between the roles of every pair of species in the network, generating a dissimilarity matrix of all species. We then performed a basic clustering analysis of the species-role dissimilarity matrix to find the most distinct groups of roles (Fig. 11). Finally, we used the function *betadisper* from the R package *vegan* (Oksanen et al., 2017) to perform the Principal Coordinates Analysis (PCoA) of the data. We found four characteristic groups of species presenting very distinct motif-role profiles. Notice that the same analysis can also be done for the motif-role profile of every link in the network. This is useful because it shows the diver-

sity of structural roles in this community and underlines how those profiles could be used to compare species, links or networks within and across ecosystems, environments and biomes (Poisot et al., 2012; Bramon Mora et al., 2018).

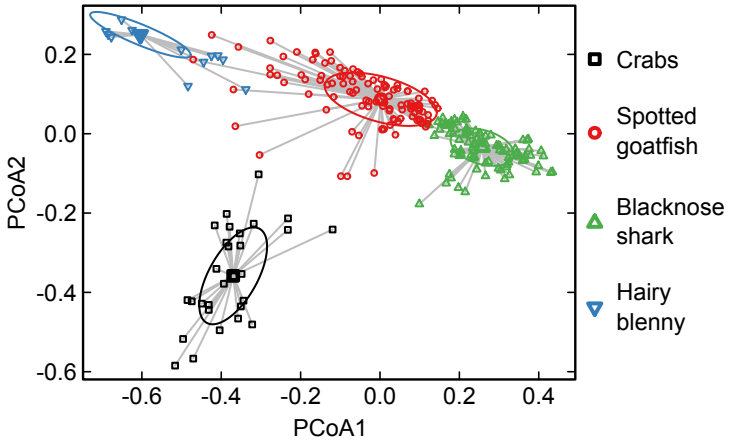


Figure 11: Principal coordinate analysis of the dissimilarity matrix containing the pairwise distances between all nodes' motif-role profiles in the marine food web under study. Every point represents a different species and each color corresponds to a group characterizing a distinct role. The species in the legend are those corresponding to the medoids of each group. The ellipses are the one-standard-deviation ellipses about the group medians.

3.5 *Availability and Future Directions*

pymfinder is open source software that can be freely and anonymously downloaded from `https://github.com/stoufferlab/pymfinder`. The documentation of the package is attached as supplementary material and the data used to test the software can be found within the github repository. *pymfinder* has been tested to run on any platform that supports Python. *pymfinder* will require you to have the Python modules Numpy and Setuptools installed in your machine. Data used to present the software has been previously published by Bascompte et. al. (Bascompte et al., 2005). We are currently working on additional software that uses the weighted motif-role profiles of nodes to efficiently align bipartite networks multiple times.

BIBLIOGRAPHY

Marti J Anderson. Distance-based tests for homogeneity of multivariate dispersions. *Biometrics*, 62(1):245–253, 2006.

Nick J Baker, Riikka Kaartinen, Tomas Roslin, and Daniel B Stouffer. Species' roles in food webs show fidelity across a highly variable oak forest. *Ecography*, 38(2):130–139, 2015.

Jordi Bascompte and Pedro Jordano. *Mutualistic networks*. Princeton University Press, 2013.

Jordi Bascompte and Carlos J Melián. Simple trophic modules for complex food webs. *Ecology*, 86(11):2868–2873, 2005.

Jordi Bascompte, Pedro Jordano, Carlos J Melián, and Jens M Olesen. The nested assembly of plant–animal mutualistic networks. *Proceedings of the National Academy of Sciences*, 100(16):9383–9387, 2003.

Jordi Bascompte, Carlos J Melián, and Enric Sala. Interaction strength combinations and the overfishing of a marine food web. *Proceedings of the National Academy of Sciences*, 102(15):5443–5447, 2005.

Jonathan J Borrelli. Selection against instability: stable subgraphs are most frequent in empirical food webs. *Oikos*, 124(12):1583–1588, 2015.

Bernat Bramon Mora, Dominique Gravel, Gilarranz J. Luis, Poisot Thimothee, and Daniel B. Stouffer. Identifying a common backbone of interactions underlying food webs from different ecosystems. *Nature Communications*, 2018. (accepted).

Juan Camacho, Daniel B Stouffer, and Luís A Nunes Amaral. Quantitative analysis of the local structure of food webs. *Journal of Theoretical Biology*, 246(2):260–268, 2007.

Alyssa R Cirtwill and Anna Eklöf. Feeding environment and other traits shape species' roles in marine food webs. *Ecology Letters*, 21 (6):875–884, 2018.

Alyssa R Cirtwill and Daniel B Stouffer. Concomitant predation on parasites is highly variable but constrains the ways in which parasites contribute to food web structure. *Journal of Animal Ecology*, 84 (3):734–744, 2015.

Alyssa R Cirtwill, Tomas Roslin, Claus Rasmussen, Jens Mogens Olesen, and Daniel B Stouffer. Between-year changes in community composition shape species' roles in an arctic plant–pollinator network. *Oikos*, 2018.

Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695(5): 1–9, 2006.

Carsten F Dormann, Jochen Fründ, Nico Blüthgen, and Bernd Gruber. Indices, graphs and null models: analyzing bipartite ecological networks. *The Open Ecology Journal*, 2(1), 2009.

Mark Emmerson and Jon M Yearsley. Weak interactions, omnivory and emergent food-web properties. *Proceedings of the Royal Society B: Biological Sciences*, 271(1537):397–405, 2004.

Roger Guimera, Marta Sales-Pardo, and Luís A Nunes Amaral. Module identification in bipartite and directed networks. *Physical Review E*, 76(3):036102, 2007.

Nadav Kashtan, Shalev Itzkovitz, Ron Milo, and Uri Alon. Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. *Bioinformatics*, 20(11):1746–1758, 2004.

Janis Klaise and Samuel Johnson. The origin of motif families in food webs. *Scientific Reports*, 7(1):16197, 2017.

Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.

Ron Milo, Nadav Kashtan, Shalev Itzkovitz, Mark EJ Newman, and Uri Alon. On the uniform generation of random graphs with prescribed degree sequences. *arXiv preprint cond-mat/0312028*, 2003.

Anje-Margriet Neutel, Johan AP Heesterbeek, and Peter C de Ruiter. Stability in real food webs: weak links in long loops. *Science*, 296 (5570):1120–1123, 2002.

Mark EJ Newman. Spread of epidemic disease on networks. *Physical Review E*, 66(1):016128, 2002.

Jari Oksanen, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs, and Helene Wagner. *vegan: Community Ecology Pack-*

age, 2017. URL https://CRAN.R-project.org/package=vegan. R package version 2.4-4.

Timothée Poisot, Elsa Canard, David Mouillot, Nicolas Mouquet, and Dominique Gravel. The dissimilarity of species interaction networks. *Ecology Letters*, 15(12):1353–1361, 2012.

María C Rodríguez-Rodríguez, Pedro Jordano, and Alfredo Valido. Functional consequences of plant-animal interactions along the mutualism-antagonism gradient. *Ecology*, 98(5):1266–1276, 2017.

James G Sanderson, Michael P Moulton, and Ralph G Selfridge. Null matrices and the analysis of species co-occurrences. *Oecologia*, 116 (1-2):275–283, 1998.

Shai S Shen-Orr, Ron Milo, Shmoolik Mangan, and Uri Alon. Network motifs in the transcriptional regulation network of escherichia coli. *Nature Genetics*, 31(1):64, 2002.

Olaf Sporns. Network analysis, complexity, and brain function. *Complexity*, 8(1):56–60, 2002.

Daniel B Stouffer, Juan Camacho, and Luís A Nunes Amaral. A robust measure of food web intervality. *Proceedings of the Royal Society B: Biological Sciences*, 103(50):19015–19020, 2006.

Daniel B Stouffer, Juan Camacho, Wenxin Jiang, and Luís A Nunes Amaral. Evidence for the existence of a robust pattern of prey selection in food webs. *Proceedings of the Royal Society B: Biological Sciences*, 274(1621):1931–1940, 2007.

Daniel B Stouffer, Marta Sales-Pardo, M Irmak Sirer, and Jordi Bascompte. Evolutionary conservation of species' roles in food webs. *Science*, 335(6075):1489–1492, 2012.

Sebastian Wernicke and Florian Rasche. Fanmod: a tool for fast network motif detection. *Bioinformatics*, 22(9):1152–1153, 2006.

Richard J Williams and Neo D Martinez. Simple rules yield complex food webs. *Nature*, 404(6774):180–183, 2000.

Esti Yeger-Lotem, Shmuel Sattath, Nadav Kashtan, Shalev Itzkovitz, Ron Milo, Ron Y Pinter, Uri Alon, and Hanah Margalit. Network motifs in integrated cellular networks of transcription–regulation and protein–protein interaction. *Proceedings of the National Academy of Sciences*, 101(16):5934–5939, 2004.

Part IV

DYNAMICS

4

UNTANGLING THE SEASONAL DYNAMICS OF PLANT-POLLINATOR COMMUNITIES

Bernat Bramon Mora[1], Eura Shin[1], Paul CaraDonna[2,3], Daniel B. Stouffer[1]

1 Centre for Integrative Ecology, School of Biological Sciences, University of Canterbury, Christchurch, New Zealand

2 Chicago Botanic Garden, Chicago, IL, USA

3 Rocky Mountain Biological Laboratory, Crested Butte, CO, USA

4.1 Abstract

A major goal in ecology is to understand the way communities vary over time. Unfortunately, we are still searching for the best approach to untangle the complexity of changes in populations and the interactions between them. An approach that has yet to be fully explored is to exhaustively compare the structure of these communities—i.e. their network of interactions—along ecological gradients or within time series. By comparing networks within a changing environment, one can understand how the roles of different species vary and change under different pressures. Here, we apply a novel network-comparison technique—network alignment—as a way to study the structure of plant-pollinator networks over time, untangling the mechanisms by which species' positions vary within these networks. We expect that these results will provide deeper understanding of the dynamics of plant-pollinator networks as well as potentially inform predictions regarding the effects of perturbations and environmental variability.

4.2 Introduction

Ecological communities are inherently dynamic. Their species composition is in constant change due to species' intrinsic phenologies (Olesen et al., 2008), affected by environmental variability (Petchey et al., 2010; Menke et al., 2012), and impacted by disturbances such as habitat fragmentation and invasive species (Aizen et al., 2008, 2012). In turn, the presence, absence, and intensity of ecological interactions also vary over time (Vázquez et al., 2007). This happens either by the direct turnover of interacting species or by higher-order effects of changes in the community composition (Poisot et al., 2015). That is,

the arrival of a new species in a community will come hand-in-hand with a new set of interactions, and these changes in the community will also indirectly interfere with existing interactions (e.g. potentially generating new cases of apparent competition between species; Holt and Kotler 1987).

Synthesizing the complex ecological dynamics observed in the field into a general framework has become one of the key challenges in ecology (Pellissier et al., 2018). Perhaps one of the main obstacles to doing so is finding the "appropriate" scale. For example, there is a longstanding tradition of exploring dynamic phenomena using mathematical models (May, 1972; Thébault and Fontaine, 2010). The empirical basis of such models is generally static networks, where nodes and links represent the different species and their observed interactions are aggregated over sampling seasons (de Ruiter et al., 2005; Ings et al., 2009; Poisot et al., 2015). However, the aforementioned dynamic nature of ecological communities makes this static scaffold unrealistic—for example, plant-pollinator communities have been shown to present high levels of within-season species turnover (Simanonok and Burkle, 2014; CaraDonna et al., 2017). On the other hand, several pioneering empirical studies have laid the groundwork for analysing natural systems over different time scales, providing crucial examples of the way ecological communities evolve within seasons (Petanidou et al., 2008; Olesen et al., 2008; Kaiser-Bunbury et al., 2009; MacLeod et al., 2016), across seasons (Alarcón et al., 2008; Ponisio et al., 2017), and over much longer time scales (Burkle et al., 2013; Cirtwill et al., 2018). These examples can often be represented as network time series, providing glimpses of ecological dynamics at whole-community scales. That said, the amount of ecological data associated to such a representation poses a new challenge. Namely, although network time series provide a more complete picture of the temporal dynamics of who interacts with whom, the comparison and synthesis of this amount of ecological data can be extremely difficult.

Several work has provided theoretical foundations to untangle the many levels of information encoded within network time series. For example, some studies have focused on the change of species composition over time, adopting a 'full-network' perspective to community dynamics. Unfortunately, while it appears useful to explain observed species distributions (Gravel et al., 2011; Dáttilo et al., 2013), these studies often need to assume that interactions are independent from local changes in species abundances, thereby washing away a key component of community dynamics. Alternatives have found success quantifying temporal interaction turnover and linking such turnover to species' phenologies (Poisot et al., 2012; CaraDonna et al., 2017). Nevertheless, these 'species-level' approaches are often cen-

tered around quantifying variation of species interactions and lack the resolution to understand how such variation transforms the overall structure of ecological networks. That is, changes in species composition or interactions might not always translate into meaningful changes in the community structure (or vice versa).

We adopt an approach here to study the seasonal dynamics of plant-pollinator communities using the technique of network alignment (Bramon Mora et al., 2018b). Conceptually, aligning ecological networks proceeds by pairing up species from different communities that play similar structural roles (Stouffer et al., 2012). This pairing essentially identifies species with analogous "positions" across communities (Fig. 12)—i.e. species that are similarly embedded in the corresponding network of interactions. It also offers a new scale to study community dynamics, one in which the state of any given species is always defined *relative* to all the other species in the community. This new scale allows us to synthesize the information encoded within network time series, providing a conceptual mapping of the changes in the communities and their components. In particular, we can use the information regarding who gets paired with whom in the alignment of the networks of a time series to ask crucial questions regarding the attachment strategies of new species (Olesen et al., 2008; Ponisio et al., 2017), the way species change their interaction partners (i.e. species' structural dynamics; CaraDonna et al. 2017) and the process in which communities lose species over a season (Burkle et al., 2013).

Leveraging network alignment, we first assess the extent to which the positions of individual plant and pollinator species are variable within seasons. That is, given the alignment between the network observed in a community at two points in time, we use the information about who gets paired with whom to reveal whether and how species change their positions over time. To understand any such changes, we then evaluate the similarity of these positions and identify the distinct groups of species' positions found across networks. This allows us to synthesize the complex dynamics of individual species over time into something much simpler: the movement of species across position groups. We finally characterize this movement, revealing the underlying structural dynamics of plants and pollinators within seasons. Overall, our study leads us towards some general rules regarding species' interaction turnover, phenology, and assembly processes in empirical plant-pollinator communities.

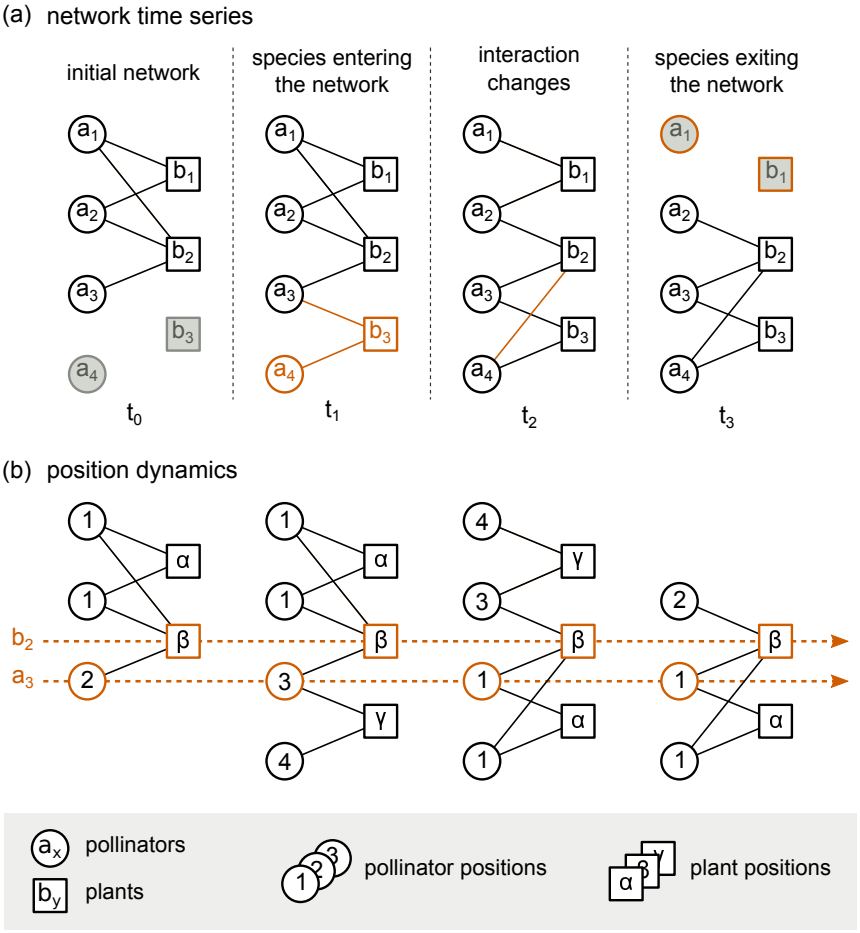


Figure 12: Dynamics of a bipartite network. (a) An example of a time series for a plant-pollinator network. The circles and squares represent pollinators and plants, respectively. The links characterize interactions between these species. The coloured species and links identify the changes made to the network over time. (b) The change in species' positions in the network time-series represented in (a). The different numbers describe different pollinator positions, and the different Greek letters describe different plant positions. The coloured dotted lines indicate the position of two specific species a_3 and b_3. On the one side, species a_3 change its position over time, starting in position "2" and ending up in position "1". On the other, species b_2 preserves the same position "β" over time. Note how several species can have the same position in the network.

4.3 *Materials and Methods*

4.3.1 *Empirical data*

We studied plant-pollinator interaction networks from a subalpine community in the Rocky Mountains (CaraDonna et al., 2017). These data were sampled at weekly intervals over three sampling seasons,

and contain nearly 30000 pairwise interactions between a total of 93 pollinator species and 46 flowering plants. To study the dynamics of these plant-pollinator communities, we aggregated the observed interactions into weekly plant-pollinator networks, where the weight of all interactions was set to the absolute number of observed interactions between the corresponding species pair during that week. In total, this resulted in three seasonal network time series comprised of 12, 15, and 16 weekly networks, respectively.

4.3.2 *Network alignment*

To analyse the dynamics of these network time series, we used the alignment technique introduced by Bramon Mora et al. (2018b). Given two networks A and B, this technique pairs up the species $i \in A$ and $j \in B$ together using the "structural roles" that they play in their respective communities. These roles are descriptors of the patterns of interactions in which species take part (Stouffer et al. 2012; Bramon Mora et al. 2018a; "Supplementary Methods" section of Appendix D). As described in Bramon Mora et al. (2018b), aligning networks is a stochastic process, where multiple random alignments λ between the species in A and B are proposed in order to find the optimal pairing between these species' roles—i.e. the optimal alignment λ^* between A and B. As a result, this optimal alignment provides us with three key pieces of information: (i) the optimal species-species pairing between all species $i \in A$ and $j \in B$; (ii) a cost function C_λ characterizing the similarity between A and B; and (iii) the contribution c_{ij} of every species-species pairing to the overall cost function. On the one side, the species-species pairing identifies species that are similarly embedded within their respective networks; since the alignment pairs up species that occupy similar positions across networks. On the other, C_λ and c_{ij} describe the "quality" of the alignment and each species-species pairing, respectively.

Notice that aligning networks can have multiple equally valid solutions. Therefore, one needs to align any pair of networks multiple times in order to properly compare their structures. For example, a given set of n alignments $\{\lambda\}$ between A and B can reveal multiple species-species parings that minimize the cost function C_λ, uncovering all pairs of species that share analogous positions across networks.

4.3.3 *Species' positions*

We first focused on the study of species' positions to understand how these change over time. To do so, we analysed (i) the uniqueness in species' positions within networks and (ii) the variability of these positions across networks. An analysis of the uniqueness of species' po-

sitions reveals how many species share the same position within any network, and it is an important measure as it unveils the symmetries in the structure of these communities. An analysis of the position variability across networks reveals instead how much the position of any given species change across networks, shedding light on the structural dynamics of species over time.

Position uniqueness within networks

To analyse whether or not species' positions were unique, we separately compared the structure of the community at each time point. If a given species i has a unique position within a network A, the alignment of A with itself should always pair up i with itself. In contrast, if i has a position that is not unique, repeating the same alignment should result in different pairings for species i—also pairing i with all species within A that share the same position. Accordingly, we aligned every weekly plant-pollinator network in our dataset to itself 100 times. This allowed us to identify the distinct pairings of species that produced optimal alignments. We then measured the uniqueness of the position of a species in a network as the proportion of alignments in which a species was paired to itself.

Position variability across networks

To measure the variability of species' positions over time, we compared networks at different time points. Specifically, we aligned every pair of networks in a given season 100 times and analysed species that are common to any of these pairs. For example, given two networks A_{t_1} and A_{t_2} that were collected at two time points t_1 and t_2, we wanted to test whether or not any species i present in both networks—i.e. a species $i \in A_{t_1} \cap A_{t_2}$—changed its position over time. To do so, we looked at whether or not species $i \in A_{t_1}$ was paired to species $i \in A_{t_2}$ in the alignment between A_{t_1} and A_{t_2}. Following this, we measured the position variability of a species i as the probability of i being paired to a different species j in any alignment between networks containing i.

We then measured how much the positions of individual species changed over time. To do so, we re-aligned the networks while artificially fixing the pairing of common species. That is, for any pair of networks A_{t_1} and A_{t_2} with n common species i, we performed n alignments in which we individually fixed the pairing of each species i and freely aligned the rest. The difference in the quality of the alignment due to the fixing of any i-i species pairing can be used to measure i's change of position. For example, if the quality of the pairing i-i is very low, it means that species i has significantly changed positions from t_1 to t_2. Similarly, the pairing i-i will be perfect if species

i presents the same position in A_{t_1} and A_{t_2}. Following this, we used this comparison across all alignments between the networks in our dataset to reveal the effects of fixing individual species' pairings, focusing on the differences between fixing plant and pollinator species.

4.3.4 *Identifying distinct groups of species' positions*

In order to further understand the changes in species' positions over time, we studied their similarity across networks. Two species $i \in A$ and $j \in B$ could have very similar positions and still not be paired in the alignment between A and B. This would happen, for example, if there was a third species $k \in B$ that also had an identical position to i—in which case species $i \in A$ would be paired to species $k \in B$. The similarity between i and j, however, can be studied by aligning the networks A and B to other networks $\{C, D, ...\}$. In doing so, we would likely observe some degree of overlap between the pairings of i and j across these other networks. This overlap would indicate that i and j share similar positions.

To uncover these types of similarities between all species' positions in our dataset, we compiled here all alignments into an alignment matrix M describing who is paired with whom within and across networks ("Supplementary Methods" section of Appendix D). In this matrix, every element m_{ij}^{AB} accounts for the pairings between any two species i and j from any two given networks A and B, respectively. Specifically, it accounts for how often these two species are paired following a set of alignments $\{\lambda\}$ between A and B as well as for the quality of such pairing. Notice that M contains information regarding all alignments between all networks in our dataset. Therefore, every row or column of this matrix represents a species k of a given network X, describing all its pairings across networks.

The alignment matrix M allowed us to analyse the similarity across all species' positions in our dataset. In particular, we focused on identifying distinct groups of species' positions that were similar within and across networks. To do so, we used a 'short random walks' algorithm to identify the modules within the matrix M describing sets of species' positions that align more often with each other than they do with the rest (Pons and Latapy, 2006). Given this grouping, we described the position groups using basic information about them: the average number of interactions (species' degree); the average degree of the best connected interacting partner; and the average degree across of interacting partners. Notice that other community detection methods can produce other valid groupings (see Supplementary Table 4). Therefore, we used a normalized mutual information analysis to study the agreement of the 'short random walks' algorithm with other community detection methods designed to analyse large

weighted undirected graphs such as the alignment matrix used here (Danon et al., 2005).

4.3.5 *Species structural dynamics across groups*

Finally, we examined the movement of pollinator species across the different network positions using a probabilistic model. Given n groups of positions, our model describes a scenario in which any species can be found in $n+2$ possible states $\vec{y}$ at time t. These include: n states $y_1 \dots y_n$ characterizing a species in each of the different groups of positions; a state $y_{n+1} = y_{pre}$ describing a species that has not yet entered the network; and a state $y_{n+2} = y_{post}$ describing a species that has already exited the network.

We used a Bayesian multinomial logistic regression as a way to estimate the rate of movement between the different states of the species over time (McElreath, 2018). In particular, our model considers $n+2$ types of events, describing the transition probability of species from any state y_i to any possible state y_j at time t as:

$$\Pr\left(y_j | s_1\left(y_i, t\right), s_2\left(y_i, t\right), ..., s_{n+2}\left(y_i, t\right),\right) = \frac{\exp s_j}{\sum_{k=1}^{n+2} \exp s_k}, \quad (9)$$

where $s_k\left(y_i, t\right)$ are *scores* that determine the resulting probabilities. To infer these scores, we first fixed one of them (e.g. assigning $s_1\left(y_i, t\right) = 0$), which serves as an arbitrary baseline. We then estimated the remaining scores as $n+1$ linear models of the form $s_k\left(y_i, t\right) = \sum_{l=1}^{n+2}\left(\alpha_{kl} + \beta_{kl} \times t\right)\delta_{il}$, where α_{kl} and β_{kl} are the parameters inferred by the model, and δ_{il} is a Kronecker delta that is set to 1 if $y_l = y_i$, and 0 otherwise. Importantly, species are assumed not to re-enter the network once they have exited it, making the transition from state y_{post} to any other state impossible. We also treated species phenologies as uninterrupted; therefore, we considered any observation of a species transition from any state y_k to y_{pre} and back to any state y_k during its activity period to be the result of a detection error. Note that we ignored any of such detection errors when inferring the probabilities.

4.4 *Results*

We first analysed the uniqueness of species positions. We found that, in general, species tend to be paired to themselves in most of the cases, indicating the presence of very few symmetric positions within any given network (Fig. 13a). That is, species are uniquely positioned at each point in time. This is true for both plants and pollinators. Importantly, the same alignments performed using unweighted networks showed the uniqueness of species positions to be much lower in this

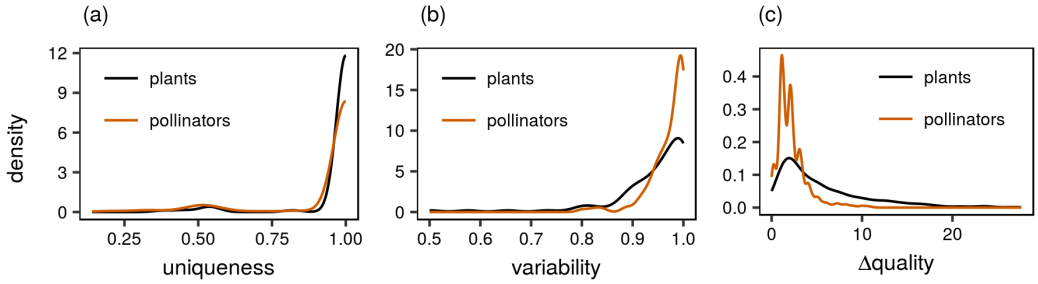


Figure 13: Analysis of species' positions across networks. (a) Uniqueness of the position of plant and pollinator species within networks. (b) Variability of species' position across networks, including only plant and pollinator species that appear in multiple networks during a season. (c) Alignment quality decrease due to the fixing of individual species that are common across networks.

type of networks (Supplementary Fig. 34). That is, we found that interaction strengths add a crucial layer of information to distinguish between species' positions in ecological networks.

Given that the positions of species within any network tend to be unique, we next studied the variability of species' position over time. We found that nearly all species that appear in multiple networks tend to change positions, presenting very high position variability from network to network (Fig. 13b). Again, this was true for both plants and pollinators. We also found this to be true when considering unweighted networks (Supplementary Fig. 34). Then, we fixed the pairing of common species over time, studying the affect of any individual species preserving its position over time. Somewhat more surprisingly, we found the fixed plant species to align worse than the fixed pollinator species. This suggests that plant species change their positions more drastically than pollinators over time (Fig. 13c).

The observed variability of the species' position over time hints at the degree of complexity that the dynamics of plant-pollinator networks encompass. As an attempt to reduce this complexity, we studied the similarity of all species' positions in our dataset using the alignment matrix M. In particular, we wanted to test whether or not there are fundamental groups of similar positions across networks. Here, we found evidence to support the idea that there are 3 distinct groups of pollinators' positions and 3 distinct groups of plants' positions (Fig. 15a). Then, we characterized the nature of these distinct groups by measuring basic node properties for each of them. Focusing on the three groups of pollinator species for the 2015 sampling season (A, B and C in Fig. 15a), we found that these show characteristic differ-

ences: group A represents species with low degree that interact with at least one generalist plant species; group B represents species with high degree that also interact with at least one generalist plant species (Fig. 15b); and, group C represents species with instead low degree and that interact with low degree plant species. These results were consistent across seasons (Supplementary Figs 35 and 36), and analogous to the results found for the plant species in the networks (Supplementary Figs 37–39). In addition, the groupings showed a strong agreement with the results found using other valid community detection methods (see Supplementary Table 4 and Supplementary Fig. 40).

Finally, we modelled the movement of the species across the different groups found in the alignment matrix using a multinomial logistic regression. In particular, for every season independently, we estimated the time-dependent transition probabilities for the species moving across groups. We found that the results showcase a consistent dynamic underlying plant-pollinator networks across seasons (Fig. 14). Given the three groups of pollinator species found across networks, this dynamic can be described as follows: the position of species entering the network tend to be from group A; once in group A, these species tend to either stay in the same group, exit the network or move to group B; species entering group B tend to either stay in the group or move back to group A; and species entering group C either move to group A, exit the network or stay in group C. Notice that similar results were found for the plant species in the network (Supplementary Fig. 41). In addition, we considered two other forms for the multinomial logistic regression, one with constant transition probabilities over time as well as another one with density-dependent transition probabilities (see "Supplementary Methods" section of Appendix D). Similar results were also found using such probabilistic models (Supplementary Figs 42–43).

4.5 *Discussion*

Our work outlines a new conceptual framework to study the dynamics of ecological communities. In particular, we focus on the idea of species' positions as a way to disentangle the seasonal dynamics of plant-pollinator networks. First, we studied the uniqueness and variability of species' positions within and across networks, respectively. We found that species have very unique positions at every time point, but they also tend to change such positions over time. Assessing the similarity of positions over time, we then found that there are major groups of positions characterizing plant-pollinator communities. These groups of positions provide a new scale to synthesizing complex ecological dynamics. Finally, we estimated the underlying

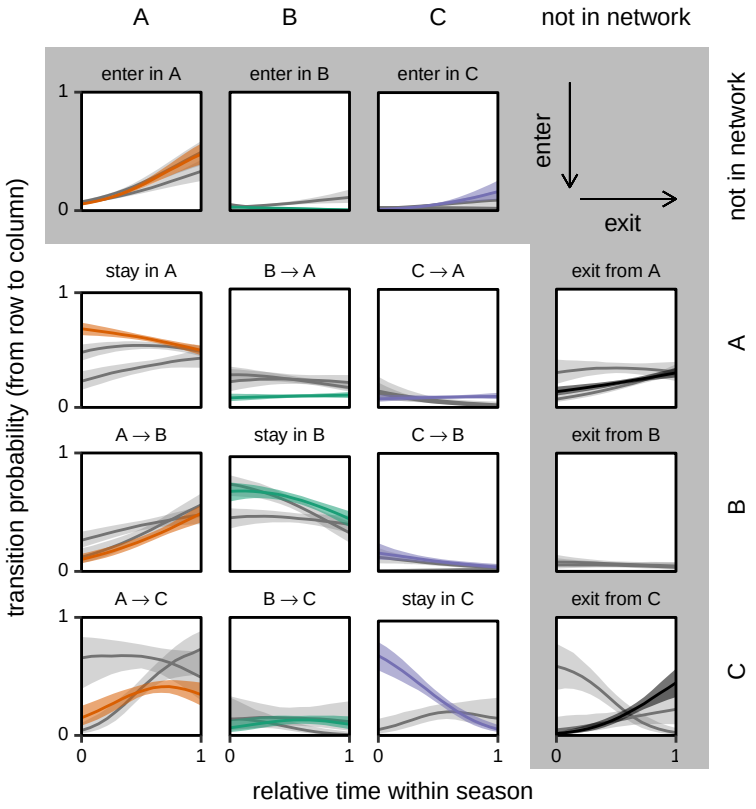


Figure 14: Movement of pollinator species across position groups. Every panel in the matrix describes the transition probability between the different groups of positions. The order of the matrix is such that it characterizes the transition probabilities from row groups to column groups over time. The different groups are those presented in Fig. 15. The shaded row describes the probabilities of species entering the network into the different groups (A, B and C) whereas the shaded column describes the probabilities of species exiting the network from each of these groups. The different lines in the graph represent the results found for every sampling year, where the shade of each line characterizes the first and third quantiles. The coloured lines indicate the results estimated for the 2015 sampling season. Notice that the color is chosen based on the recipient group of the transition probability.

dynamics of species pollinators within seasons and found general rules regarding species' seasonal dynamics within plant-pollinator communities. Putting this all together, our results suggest that the structure of plant-pollinator networks is extremely dynamic, where species rapidly switch positions within the network over a season. This structural dynamic, however, is also coherent across years, and one can predict the changes in species' positions within networks over time.

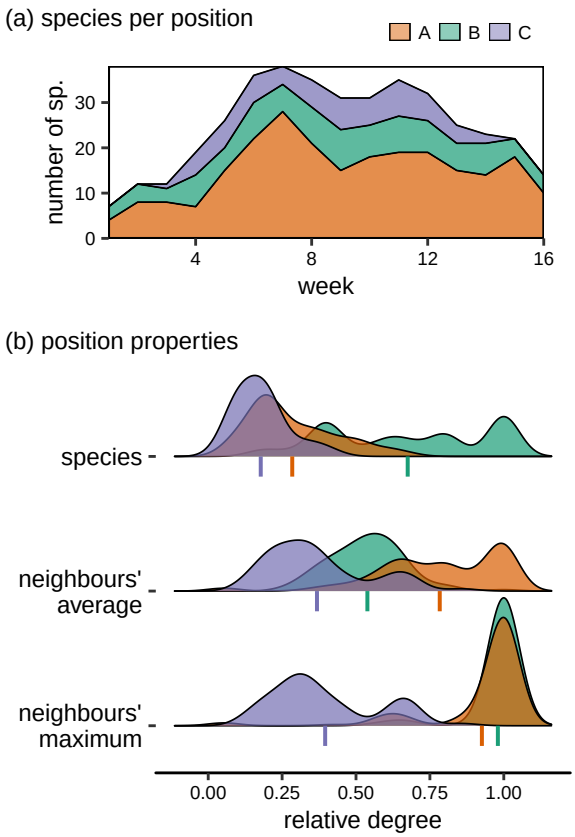


Figure 15: Distinguishing properties of the distinct groups of pollinator positions found for the 2015 sampling season. (a) Number of pollinator species in each group of positions over time. Each color represents a different group of positions. (b) Different properties summarizing the species forming each group. The top panel shows the relative degree distribution of species in each group. The middle panel shows the neighbours' average relative degree for the species in each group. The bottom panel depicts the relative degree distribution of the best connected neighbour of every species in each group. The coloured segments depicted under the distributions characterize the mean of each distribution.

The study of network time series is challenging due to the many levels of information that these systems encompass. One could, for example, adopt a full-network perspective and study community dynamics using general network metrics (Dunne et al., 2002). Unfortunately, network metrics lack the resolution to distill the mechanisms by which species change positions over time (Ponisio et al., 2017). Indeed, the study of metrics such as nestedness and connectance has shown certain mutualistic networks to exhibit generally constant structures over time (Bascompte et al., 2003; Petanidou et al., 2008; Chacoff et al., 2018). While this may be useful for understanding their dynamical stability and functioning (Bastolla et al., 2009; Thébault and

Fontaine, 2010; Thompson et al., 2012), these metrics are particularly ill-suited to understand the full scope of plant-pollinator seasonal dynamics. Alternatively, one could use single-species approaches. Ecological data, however, is often clouded by environmental variability (Tylianakis et al., 2008; Garibaldi et al., 2011) or sampling errors in the data collection (Goldwasser and Roughgarden, 1997), both of which can add considerable noise to single-species dynamics. Perhaps most importantly, these approaches could also easily be overwhelmed by species' natural idiosyncrasies (Westphal et al., 2003), which could mask potential general rules governing community dynamics. Indeed, we observed the effects of such idiosyncrasies when studying the uniqueness and variability of species' positions. The high uniqueness of species' positions indicates how singularly different species are embedded within a network; and, the high variability shows how sensitive these positions are to changes in the network structure. Noticeably, our observations on the variability of species positions also agree with recent work showcasing constant temporal switching of species' interactions in other empirical plant-pollinator communities (Miele et al., 2019).

Here, we showed how it is possible to find a useful middle ground between full-network and single-species approaches. In particular, we focused on identifying distinct groups of positions within networks by clustering species with similar positions. This group scale allowed us to strategically prune down plant-pollinator dynamics. Assessing the movement of species across these groups of positions, one could, for example, focus on how pollinator species enter the community. As expected, we found the degree of newcomers to be generally low; we observe species entering the community mostly as specialists (groups A and C from Fig. 15). From these new-coming species, most of them tend to interact with at least one generalist plant, showing consistency with the idea of preferential attachment (group A; Olesen et al. 2008). Likewise, the detachment of pollinators from networks often comes from groups of less-connected species (groups A and C from Fig. 15), also in agreement with the idea of preferential detachment (Aizen et al., 2012; Burkle et al., 2013). The symmetry between these two processes—preferential attachment and detachment—have been showcased at longer time scales (Tylianakis et al., 2018) as well as hypothesised to generate and maintain network patterns promoting stability (Bascompte and Stouffer, 2009). More generally, our results display a road map on how species change positions within the community, what positions are the ones species take before exiting the network, and which species will likely stay in the network the longest. Importantly, we found these dynamics independently replicated across the three different sampling seasons comprised in the dataset. Moreover, different community detection methods can par-

tition the alignment matrix differently, providing different degrees of resolution to the dynamics of species across groups. Although, we showed how more complex partitions display more resolved dynamics (Supplementary Fig. 40), we also found that finer resolutions might lead to groups of positions that can be difficult to discern from each other (Supplementary Fig. 44).

Among other potential implications of our results, we believe that they could be crucial to designing mathematical models that can simulate the seasonal dynamics of plant-pollinator communities. An important body of work has been produced on the study of stability and persistence to perturbations of plant-pollinator networks (Fortuna and Bascompte, 2006; Thébault and Fontaine, 2010; Saavedra et al., 2011). Unfortunately, with few exceptions (Kaiser-Bunbury et al., 2010; Saavedra et al., 2016), most such work is often built around mathematical models that consider ecological communities as static entities (Rohr et al., 2014), ignoring the intrinsic phenologies of species forming the communities. Following the observed variability of species positions within networks and species' coherent dynamics, it is difficult to overlook the fact that such dynamical models are a broad approximation of what empirical plant-pollinator dynamics actually look like. We suggest designing models that reflect the changing nature of plant-pollinator interactions, potentially using the groups of positions as the building blocks for such models (similar to the concept of trophic components in food web; Bascompte and Melián 2005; Kondoh 2008). For example, one could consider the use of time-dependent stochastic block models (Xu and Hero, 2014) and simplify the overall dynamics of ecological communities using the groups of positions as the representative nodes in a network between pollinator and plant groups.

Finally, we identify three areas we feel represent key steps from which to move forward. First, the approach used in the present work is not limited to plant-pollinator networks. Indeed it could be used to shed light on the mechanisms governing many other systems, including food webs (Hart et al., 2000), host-parasite communities (Pilosof et al., 2013) or other types of temporal networks (Holme and Saramäki, 2012). Though we focused on temporal variation, another interesting perspective would be to put the same tools to work across other type of gradients (Tylianakis and Morris, 2017). For example, one could focus on the structural variability of plant-frugivore networks along forest-farmland gradients (Menke et al., 2012), which could reveal how bird species change positions within networks in order to adapt to different environmental conditions. In addition, we here defined species' position purely is based on the structure of plant-pollinator communities. Nevertheless, these positions could easily also account

for other species' properties such as species' ecological traits and evolutionary histories (Bramon Mora et al., 2018b). This would allow us, for example, to study network dynamics from a functional diversity or evolutionary perspective, potentially untangling the eco-evolutionary mechanisms governing complex community dynamics.

BIBLIOGRAPHY

Marcelo A Aizen, Carolina L Morales, and Juan M Morales. Invasive mutualists erode native pollination webs. *PLoS Biology*, 6(2):e31, 2008.

Marcelo A Aizen, Malena Sabatino, and Jason M Tylianakis. Specialization and rarity predict nonrandom loss of interactions from mutualist networks. *Science*, 335(6075):1486–1489, 2012.

Ruben Alarcón, Nickolas M Waser, and Jeff Ollerton. Year-to-year variation in the topology of a plant–pollinator interaction network. *Oikos*, 117(12):1796–1807, 2008.

Jordi Bascompte and Carlos J Melián. Simple trophic modules for complex food webs. *Ecology*, 86(11):2868–2873, 2005.

Jordi Bascompte and Daniel B Stouffer. The assembly and disassembly of ecological networks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1524):1781–1787, 2009.

Jordi Bascompte, Pedro Jordano, Carlos J Melián, and Jens M Olesen. The nested assembly of plant–animal mutualistic networks. *Proceedings of the National Academy of Sciences*, 100(16):9383–9387, 2003.

Ugo Bastolla, Miguel A Fortuna, Alberto Pascual-García, Antonio Ferrera, Bartolo Luque, and Jordi Bascompte. The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature*, 458(7241):1018, 2009.

Bernat Bramon Mora, Alyssa R Cirtwill, and Daniel B Stouffer. pymfinder: a tool for the motif analysis of binary and quantitative complex networks. *bioRxiv*, page 364703, 2018a.

Bernat Bramon Mora, Dominique Gravel, Luis J Gilarranz, Timothée Poisot, and Daniel B Stouffer. Identifying a common backbone of interactions underlying food webs from different ecosystems. *Nature Communications*, 9(1):2603, 2018b.

Laura A Burkle, John C Marlin, and Tiffany M Knight. Plant-pollinator interactions over 120 years: loss of species, co-occurrence, and function. *Science*, 339(6127):1611–1615, 2013.

Paul J CaraDonna, William K Petry, Ross M Brennan, James L Cunningham, Judith L Bronstein, Nickolas M Waser, and Nathan J Sanders. Interaction rewiring and the rapid turnover of plant–pollinator networks. *Ecology Letters*, 20(3):385–394, 2017.

Natacha P Chacoff, Julian Resasco, and Diego P Vázquez. Interaction frequency, network position, and the temporal persistence of interactions in a plant–pollinator network. *Ecology*, 99(1):21–28, 2018.

Alyssa R Cirtwill, Tomas Roslin, Claus Rasmussen, Jens Mogens Olesen, and Daniel B Stouffer. Between-year changes in community composition shape species' roles in an arctic plant–pollinator network. *Oikos*, 127(8):1163–1176, 2018.

Leon Danon, Albert Diaz-Guilera, Jordi Duch, and Alex Arenas. Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(09):P09008, 2005.

Wesley Dáttilo, Paulo R Guimarães, and Thiago J Izzo. Spatial structure of ant–plant mutualistic networks. *Oikos*, 122(11):1643–1648, 2013.

Peter C de Ruiter, Volkmar Wolters, John C Moore, and Kirk O Winemiller. Food web ecology: playing jenga and beyond. *Science*, 309 (5731):68–71, 2005.

Jennifer A Dunne, Richard J Williams, and Neo D Martinez. Food-web structure and network theory: the role of connectance and size. *Proceedings of the National Academy of Sciences*, 99(20):12917–12922, 2002.

Miguel A Fortuna and Jordi Bascompte. Habitat loss and the structure of plant–animal mutualistic networks. *Ecology Letters*, 9(3): 281–286, 2006.

Lucas A Garibaldi, Ingolf Steffan-Dewenter, Claire Kremen, Juan M Morales, Riccardo Bommarco, Saul A Cunningham, Luísa G Carvalheiro, Natacha P Chacoff, Jan H Dudenhöffer, Sarah S Greenleaf, et al. Stability of pollination services decreases with isolation from natural areas despite honey bee visits. *Ecology Letters*, 14(10):1062–1072, 2011.

Lloyd Goldwasser and Jonathan Roughgarden. Sampling effects and the estimation of food-web properties. *Ecology*, 78(1):41–54, 1997.

Dominique Gravel, Francois Massol, Elsa Canard, David Mouillot, and Nicolas Mouquet. Trophic theory of island biogeography. *Ecol-*

ogy Letters, 14(10):1010–1016, 2011.

Deborah R Hart, Lewi Stone, and Tom Berman. Seasonal dynamics of the lake kinneret food web: the importance of the microbial loop. *Limnology and Oceanography*, 45(2):350–361, 2000.

Petter Holme and Jari Saramäki. Temporal networks. *Physics Reports*, 519(3):97–125, 2012.

Robert D Holt and Burt P Kotler. Short-term apparent competition. *The American Naturalist*, 130(3):412–430, 1987.

Thomas C Ings, José M Montoya, Jordi Bascompte, Nico Blüthgen, Lee Brown, Carsten F Dormann, François Edwards, David Figueroa, Ute Jacob, J Iwan Jones, et al. Ecological networks–beyond food webs. *Journal of Animal Ecology*, 78(1):253–269, 2009.

Christopher N Kaiser-Bunbury, Jane Memmott, and Christine B Müller. Community structure of pollination webs of mauritian heathland habitats. *Perspectives in Plant Ecology, Evolution and Systematics*, 11(4):241–254, 2009.

Christopher N Kaiser-Bunbury, Stefanie Muff, Jane Memmott, Christine B Müller, and Amedeo Caflisch. The robustness of pollination networks to the loss of species and interactions: a quantitative approach incorporating pollinator behaviour. *Ecology Letters*, 13(4): 442–452, 2010.

Michio Kondoh. Building trophic modules into a persistent food web. *Proceedings of the National Academy of Sciences*, 105(43):16631–16635, 2008.

Molly MacLeod, Mark A Genung, John S Ascher, and Rachael Winfree. Measuring partner choice in plant–pollinator networks: using null models to separate rewiring and fidelity from chance. *Ecology*, 97(11):2925–2931, 2016.

Robert M May. Will a large complex system be stable? *Nature*, 238 (5364):413, 1972.

Richard McElreath. *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC, 2018. pp. 323–326.

Sebastian Menke, Katrin Böhning-Gaese, and Matthias Schleuning. Plant–frugivore networks are less specialized and more robust at forest–farmland edges than in the interior of a tropical forest. *Oikos*, 121(10):1553–1566, 2012.

Vincent Miele, Rodrigo Ramos-Jiliberto, and Diego P Vasquez. Temporal switching of species roles in a plant-pollinator network. *bioRxiv*, page 543637, 2019.

Jens M Olesen, Jordi Bascompte, Heidi Elberling, and Pedro Jordano. Temporal dynamics in a pollination network. *Ecology*, 89(6):1573–1582, 2008.

Loïc Pellissier, Camille Albouy, Jordi Bascompte, Nina Farwig, Catherine Graham, Michel Loreau, Maria Alejandra Maglianesi, Carlos J Melián, Camille Pitteloud, Tomas Roslin, et al. Comparing species interaction networks along environmental gradients. *Biological Reviews*, 93(2):785–800, 2018.

Theodora Petanidou, Athanasios S Kallimanis, Joseph Tzanopoulos, Stefanos P Sgardelis, and John D Pantis. Long-term observation of a pollination network: fluctuation in species and interactions, relative invariance of network structure and implications for estimates of specialization. *Ecology Letters*, 11(6):564–575, 2008.

Owen L Petchey, Ulrich Brose, and Björn C Rall. Predicting the effects of temperature on food web connectance. *Philosophical Transactions of the Royal Society B*, 365(1549):2081–2091, 2010.

Shai Pilosof, Miguel A Fortuna, Maxim V Vinarski, Natalia P Korallo-Vinarskaya, and Boris R Krasnov. Temporal dynamics of direct reciprocal and indirect effects in a host–parasite network. *Journal of Animal Ecology*, 82(5):987–996, 2013.

Timothée Poisot, Elsa Canard, David Mouillot, Nicolas Mouquet, and Dominique Gravel. The dissimilarity of species interaction networks. *Ecology Letters*, 15(12):1353–1361, 2012.

Timothée Poisot, Daniel B Stouffer, and Dominique Gravel. Beyond species: why ecological interaction networks vary through space and time. *Oikos*, 124(3):243–251, 2015.

Lauren C Ponisio, Marilia P Gaiarsa, and Claire Kremen. Opportunistic attachment assembles plant–pollinator networks. *Ecology Letters*, 20(10):1261–1272, 2017.

Pascal Pons and Matthieu Latapy. Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications*, 10(2):191–218, 2006.

Rudolf P Rohr, Serguei Saavedra, and Jordi Bascompte. On the structural stability of mutualistic systems. *Science*, 345(6195):1253497,

2014.

Serguei Saavedra, Daniel B Stouffer, Brian Uzzi, and Jordi Bascompte. Strong contributors to network persistence are the most vulnerable to extinction. *Nature*, 478(7368):233–235, 2011.

Serguei Saavedra, Rudolf Rohr, Miguel Fortuna, Nuria Selva, and Jordi Bascompte. Seasonal species interactions minimize the impact of species turnover on the likelihood of community persistence. *Ecology*, 2016.

Michael P Simanonok and Laura A Burkle. Partitioning interaction turnover among alpine pollination networks: spatial, temporal, and environmental patterns. *Ecosphere*, 5(11):1–17, 2014.

Daniel B Stouffer, Marta Sales-Pardo, M Irmak Sirer, and Jordi Bascompte. Evolutionary conservation of species' roles in food webs. *Science*, 335(6075):1489–1492, 2012.

Elisa Thébault and Colin Fontaine. Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science*, 329(5993):853–856, 2010.

Ross M Thompson, Ulrich Brose, Jennifer A Dunne, Robert O Hall Jr, Sally Hladyz, Roger L Kitching, Neo D Martinez, Heidi Rantala, Tamara N Romanuk, Daniel B Stouffer, et al. Food webs: reconciling the structure and function of biodiversity. *Trends in ecology & evolution*, 27(12):689–697, 2012.

Jason M Tylianakis and Rebecca J Morris. Ecological networks across environmental gradients. *Annual Review of Ecology, Evolution, and Systematics*, 48, 2017.

Jason M Tylianakis, Raphael K Didham, Jordi Bascompte, and David A Wardle. Global change and species interactions in terrestrial ecosystems. *Ecology Letters*, 11(12):1351–1363, 2008.

Jason M Tylianakis, Laura B Martínez-García, Sarah J Richardson, Duane A Peltzer, and Ian A Dickie. Symmetric assembly and disassembly processes in an ecological network. *Ecology letters*, 21(6): 896–904, 2018.

Diego P Vázquez, Carlos J Melián, Neal M Williams, Nico Blüthgen, Boris R Krasnov, and Robert Poulin. Species abundance and asymmetric interaction strength in ecological networks. *Oikos*, 116(7): 1120–1127, 2007.

Catrin Westphal, Ingolf Steffan-Dewenter, and Teja Tscharntke. Mass flowering crops enhance pollinator densities at a landscape scale. *Ecology Letters*, 6(11):961–965, 2003.

Kevin S Xu and Alfred O Hero. Dynamic stochastic blockmodels for time-evolving social networks. *IEEE Journal of Selected Topics in Signal Processing*, 8(4):552–562, 2014.

Part V

GENERAL DISCUSSION

STRUCTURE AND DYNAMICS OF ECOLOGICAL NETWORKS

Over the course of this thesis, my co-authors and I developed multiple statistical tools to study the structure and dynamics of ecological networks. Importantly, there are already multiple methods with which ecologists have tried to shed light on the structure of ecological networks. The tools presented here were not designed to simply extend this already extensive list of methods. Instead, they were designed as investigative tools that targeted specific ecological hypothesis, answering key ecological questions regarding the structure and dynamics of ecological networks. In the following pages, I will provide an overall discussion of my work, which I will separate in three parts: a summary of results, an overview of the general implications of my work, and a description of the steps that I identify as key moving forward.

Summary of results

In the first chapter of this thesis, my co-authors and I developed a correlation-informed null model—a model that combines the classic concept of a null model (Gotelli, 2000) and the ideas underlying joint modeling (Warton et al., 2015)—to untangle the mechanisms shaping the structure of biological communities. With this model, we assessed how predictive the information encoded within different correlation matrix is for explaining network metrics that are commonly used in community ecology (Stouffer et al., 2007; Patterson and Atmar, 1986). We first focused on the study the structure of food webs, finding a strong phylogenetic component underlying food-web motifs. In particular, we found that a null model accounting for the phylogenetic relationships in predators' diets can partially explain the motif representation observed in food web, showing a prey-selection mechanism shaping the structure of these networks (Stouffer et al., 2007; Naisbit et al., 2012). Then, we investigated the factors explaining the nested pattern observed in species assemblages. We found this pattern to be predominantly explained by island species richness as opposed to island size, isolation or species range. This suggests that nestedness could potentially just be an artifact of island species richness that becomes tautological when one controls for it.

In the second chapter of this thesis, my co-authors and I developed a network alignment technique to study the structure of food webs across ecosystems. First, we tested whether or not food webs from

multiple ecosystems have significantly different structures, finding differences for almost every pair of ecosystem types. Importantly, such variability is somewhat in contrast with the strong evidence suggesting that food web share a common set of macroscopic properties across environments (Riede et al., 2010; Eklöf et al., 2013). To reconcile this two ideas, we proposed the existence of a backbone of interactions underlying all networks that is surrounded by a peripheral structure that instead differ in characteristic ways. Testing this idea using a large dataset of empirical food webs, we found the existence of two backbone structures present across all networks and across ecosystems, which could have crucial implications to understand food-web dynamics and the mechanisms by which food webs from different environments are shaped under different pressures.

In the third chapter of this thesis, my co-authors and I developed a computational tool to characterize the structural role of species in weighted networks—networks in which the links are weighted based on the interaction strengths between species. The main purpose of developing this tool was to be able to more efficiently compare bipartite networks using the method described in the second chapter. Interaction strengths can add a crucial layer of information to networks such as the ones formed by the mutualistic interactions between plant and pollinator species (McCann et al. 1998; Kokkoris et al. 1999; Bascompte et al. 2006; see also Fig. 13 and Supplementary Fig. 34). Therefore, the development of this tool was a necessary step for us to be able to properly align this type of network.

In the fourth chapter of this thesis, my co-authors and I developed a new conceptual framework to study network time series by combining the tools presented in the second and third chapters of this thesis. Focusing on the seasonal dynamics of plant-pollinator networks, this framework allowed us to shed light on the way species change positions over time. First, we studied the uniqueness of species' positions within networks, finding that species tend to have very unique positions at each point in time. However, when studying the variability of these positions across networks, we found that species also tend to change their positions over time. In order to synthesize the underlaying dynamics associated to this variability, we identified groups of species across networks that share similar positions. Given these major groups of positions, we finally studied the movement of species across these groups, untangling the general rules governing species' seasonal dynamics within plant-pollinator communities. In particular, our results displayed a road map on how species change positions within a given community, what positions are the ones species take before exiting the corresponding ecological network, and what species will likely stay in the network the longest.

Implications

The implications of the results found in each individual chapter have been discussed within the chapters themselves. Here, I will therefore discuss the impact of this thesis as a whole, starting with the link between ecological networks and evolution. Although it is unquestioned that the network approach provides a simple way to synthesize the innate complexity underlaying ecological systems, this approach unfortunately neglects the fact that species are part of a hierarchically structured phylogeny. It is widely accepted that species' interactions are determined by their phenotypic traits (Stang et al., 2009; Naisbit et al., 2012; Gravel et al., 2013), which in turn are largely inherited from species' ancestors (Wiens and Graham, 2005; Gilbert and Webb, 2007). Therefore, it is a long hold assumption in ecology that species interactions are themselves phylogenetically conserved (Bersier and Kehrli, 2008; Gómez et al., 2010; Krasnov et al., 2012), and they should not be considered independently.

The phylogenetic component of ecological networks is something that is only explicitly considered in first chapter of the thesis, shaping the patterns of interactions in stream food webs. However, this phylogenetic component certainly has implications for the results found throughout this thesis. For example, it is reasonable to think that there are potential evolutionary mechanisms driving the existence of the backbones of interactions found in the second chapter (Rezende et al., 2009; Cagnolo et al., 2011). Likewise, the structural roles studied in the third chapter have been shown to be evolutionary conserved (Stouffer et al., 2012), and the strength of the interactions could similarly encompass a phylogenetic component (Emmerson and Raffaelli, 2004). Finally, species' evolutionary histories can explain to some extend the structure of plant-pollinator communities (Hutchinson et al., 2017), and one would expect the seasonal dynamics observed in the fourth chapter to be similarly influenced.

The link between ecological networks and evolution is closely related to another central topic in network ecology: the assembly process (Bascompte and Stouffer, 2009). The way communities assemble over time and the origin of the non-random structure of networks have been well studied over the recent years (Simon, 1955; Barabási and Albert, 1999). In particular, ecologists have tried to untangle the eco-evolutionary mechanisms by which ecological networks are formed using different strategies. On the one side, multiple theoretical models of community assembly have been proposed that attempt to replicate the assembly process of antagonistic (Bastolla et al., 2005) and mutualistic (Campbell et al., 2011) networks. On the other, a long list of experimental studies have addressed the same question by analyzing large collections of data, from studying the chronological forma-

tion of ecological networks in retreating glaciers (Albrecht et al., 2010) to comparing the communities over long periods of time (Burkle et al., 2013). The results presented in this thesis also closely resonate with the formation of ecological networks. For instance, in the first chapter, our phylogenetically-informed null models showcase how prey selection mechanisms better explain food web structure than the analogous predator selection mechanisms (Stouffer et al., 2007). Further studies on the trait composition of the backbone of interactions found in the second chapter of this thesis could also be key to shed light fundamental aspects of community assembly, allowing us to understand the role of the backbone as a driver of species' coexistence and diversification (Emerson and Gillespie, 2008). Finally, the seasonal dynamics found in the fourth chapter are likewise relevant to community assembly, showing consistency with the idea of preferential attachment (Olesen et al., 2008) as opposed to other mechanisms of plant-pollinator assembly (Ponisio et al., 2017).

Among the many other potential implications of the results presented throughout this thesis, I will finally focus on the link between ecological networks and global change. Global change is currently the major threat to biodiversity and community stability (Sala et al., 2000). Species invasions, habitat fragmentation, environmental perturbations, etc. affect directly or indirectly species' abundance, physiology and interactions (Sala et al., 2000). As a result, ecological networks are experiencing critical changes that threaten the survival of entire communities (Tylianakis et al., 2008). There are a few ways in which ecologists can attempt to fully understand how ecological networks might suffer the consequences of global change; a crucial one is by means of network comparison. This comparison can take many forms. As shown in the first chapter of the thesis, one can compare ecological communities against models or null expectations in order to reveal the influence of different factors on their structure. Likewise, the comparison of networks across ecosystems can be used to understand how interactions are shaped across environments. In the second chapter of the thesis, for example, the comparison of food webs across ecosystems revealed not only a common backbone of interactions but also a differing peripheral structure. Such periphery could encode, therefore, the effects of different environments on the structure of ecological networks.

Next steps

Following the work presented here, there are several areas of research that I believe could be of great interest moving forward. One of them is predicting ecological interactions. Ecological networks are naturally variable due to factors such as environmental conditions,

dispersal limitations, and historical contingencies (Currie et al., 2004; Thompson and Townsend, 2005; Petchey et al., 2010). That said, these networks often present non-random structures (Dunne et al., 2002a; Stouffer et al., 2006; Beckerman et al., 2006; Stouffer et al., 2007; Petchey et al., 2008) that have been shown to be crucial to understand multiple aspects regarding the dynamical stability of ecological communities (Dunne et al., 2002b; Saavedra et al., 2011; Gravel et al., 2011; Allesina and Tang, 2012). Unfortunately, ecological data is often scarce, and differences in the sampling methods used to collect these data can introduce a lot of variability to the observed network structure (Magurran and McGill, 2011)—for example, some techniques have been shown to often miss the observation of weak links (Goldwasser and Roughgarden, 1997). As a result, predicting interactions has become of ecological interest over the recent years (Jordano, 2016) and multiple studies have been developed in this direction (Clauset et al., 2008; Guimerà and Sales-Pardo, 2009; Gravel et al., 2013; Eklöf et al., 2013).

The tools developed in the first chapter of this thesis could be put to work on this topic. In particular, we calculate the swapping probabilities of the correlation-informed null models by means of a Generalized Linear Mixed Model (Pearse et al. 2015; Eq. 1). Similar to the work of Olival et al. (2017) and Dallas et al. (2017), the same framework can be used to simply estimate of the probability to observe any given link in an incidence matrix. These tools could actually be extended to consider multiple random effects in Eqn (1), incorporating the information of different species traits (e.g. body size, phylogenetic relationships, etc.), as a way to predict multiple types of ecological interactions (e.g. predator-prey, plant-pollinator, host-parasitoid, seed-dispersal, etc.).

A very different approach could also be taken using the tool developed in the second chapter of this thesis—the alignment technique. One could predict "expected" interactions by mapping the networks of a large dataset on top of each other and studying their resultant interaction overlap. As we did to uncover the backbones of interactions (Fig. 4c), the link overlap would determine the probability of any two species interacting. As explained throughout the thesis, we align networks by pairing their components based on the roles that these play in their respective systems, where the definition of a role can describe any property of the nodes' nature. Therefore, the alignments could contain information regarding species' traits, abundances or phylogenetic histories, which would allow us to predict interactions considering multiple levels of information.

I find the idea of aligning networks using different definitions of a species' role particularly interesting because it relates to concepts

such as functional diversity. Functional diversity is a key component of biodiversity that describes the variety in species traits (Tilman, 1999). More importantly, it reflects the differences in species' adaptations to the environment relative to all other species (Petchey and Gaston, 2002; Dehling et al., 2016), and it is crucial to understand the assembly of ecological networks (Rezende et al., 2007, 2009). For example, the matching of species' functional traits has been shown to be a key mechanism regulating interactions in plant-animal mutualistic networks (Albrecht et al., 2018). In this context, the alignment of two networks using species' traits is conceptually equivalent to aligning these communities in the functional trait space (Dehling and Stouffer, 2018). The pairing of two specific species in the alignment between any pair of networks would reveal analogous species in the functional trait space. Following this, the identification of a backbone of interactions (second chapter) or the study of community dynamics (fourth chapter) in the light of this new 'trait alignment' could acquire a whole new dimension.

Another concept that I have not been able to fully explore here is the idea of network alignment as a management tool. I believe that the link between structure and dynamics—the one that I describe in the last project of this thesis—is especially interesting because comparing the structure of complex networks is much easier than elucidating their dynamics. Imagine two nodes that are similarly embedded in different networks. If the dynamics of these nodes is governed by the same rules, it is reasonable to think that they are likely to share the same fate. Network alignment can be used to link these nodes together. This is important because the identification of species critically affected by environmental perturbations in one community (Doak et al., 2008) could be used as a proxy for targeting other species from different communities that might be sensitive to similar disturbances (Saavedra et al., 2011). Therefore, we could use alignment to extend the empirical observation of the dynamics of one ecological network to other networks that share similar characteristics.

Finally, there are many other aspects of the relationship between structure and dynamics of ecological networks that remain unexplored. One of these aspects is the so-called structural redundancy (Clarke and Warwick, 1998). The concept of structural redundancy in ecology defines a simple idea: if one part of an ecological network fails, others will bear the load of the failing part and the entire network will persist. This idea finds parallels in other systems. In financial markets, for example, risk is often reduced by investing in a diverse portfolio (Kogut, 1994; Uzzi, 1997). Similarly, in ecology, redundancy could be proposed to be a key factor contributing to the persistence of ecosystems (Stouffer and Bascompte, 2010), hypoth-

esizing that ecological communities might have evolved to present redundant structures to prevent ecological collapse. Network alignment could add an interesting perspective to this idea since it could be used to quantify the structural redundancy of different ecological networks. Similar to the idea of uniqueness of species' positions defined in the fourth chapter of this thesis (Fig. 13), structural redundancy could quantify the within-network symmetries found in ecological communities.

Conclusions

Paraphrasing Hill (1973): now ecological networks are of theoretical interest because they can be related to community stability, biodiversity, ecosystem functioning, conservation, and global change; but, networks are mere numbers and should be distinguished from the theories which they support. I believe that the body of work presented in this thesis speaks to the heart of this idea. Networks are useful objects that my co-authors and I used in different ways to answer key ecological questions. Despite the investigative purpose with which we designed the tools presented here, however, I also believe that this thesis opens more questions than it offers answers. As shown in each chapter and the discussion above, these tools can be applied to investigate many other relevant ecological hypotheses, and I am certain that I and other researchers will benefit from them in the future.

BIBLIOGRAPHY

Jörg Albrecht, Alice Classen, Maximilian GR Vollstädt, Antonia Mayr, Neduvoto P Mollel, David Schellenberger Costa, Hamadi I Dulle, Markus Fischer, Andreas Hemp, Kim M Howell, et al. Plant and animal functional diversity drive mutualistic network assembly across an elevational gradient. *Nature Communications*, 9, 2018.

Matthias Albrecht, Matthias Riesen, and Bernhard Schmid. Plant–pollinator network assembly along the chronosequence of a glacier foreland. *Oikos*, 119(10):1610–1624, 2010.

Stefano Allesina and Si Tang. Stability criteria for complex ecosystems. *Nature*, 483(7388):205–208, 2012.

Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

Jordi Bascompte and Daniel B Stouffer. The assembly and disassembly of ecological networks. *Philosophical Transactions of the Royal Society B*, 364(1524):1781–1787, 2009.

Jordi Bascompte, Pedro Jordano, and Jens M Olesen. Asymmetric coevolutionary networks facilitate biodiversity maintenance. *Science*, 312(5772):431–433, 2006.

Ugo Bastolla, Michael Lässig, Susanna C Manrubia, and Angelo Valleriani. Biodiversity in model ecosystems, ii: species assembly and food web structure. *Journal of Theoretical Biology*, 235(4):531–539, 2005.

Andrew P Beckerman, Owen L Petchey, and Philip H Warren. Foraging biology predicts food web complexity. *Proceedings of the National Academy of Sciences*, 103(37):13745–13749, 2006.

Louis-Félix Bersier and Patrik Kehrli. The signature of phylogenetic constraints on food-web structure. *Ecological Complexity*, 5(2):132–139, 2008.

Laura A Burkle, John C Marlin, and Tiffany M Knight. Plant-pollinator interactions over 120 years: loss of species, co-occurrence, and function. *Science*, 339(6127):1611–1615, 2013.

Luciano Cagnolo, Adriana Salvo, and Graciela Valladares. Network topology: patterns and mechanisms in plant-herbivore and host-parasitoid food webs. *Journal of Animal Ecology*, 80(2):342–351, 2011.

Colin Campbell, Suann Yang, Réka Albert, and Katriona Shea. A network model for plant–pollinator community assembly. *Proceedings of the National Academy of Sciences*, 108(1):197–202, 2011.

KR Clarke and RM Warwick. Quantifying structural redundancy in ecological communities. *Oecologia*, 113(2):278–289, 1998.

Aaron Clauset, Cristopher Moore, and Mark EJ Newman. Hierarchical structure and the prediction of missing links in networks. *Nature*, 453(7191):98, 2008.

David J Currie, Gary G Mittelbach, Howard V Cornell, Richard Field, Jean-Francois Guégan, Bradford A Hawkins, Dawn M Kaufman, Jeremy T Kerr, Thierry Oberdorff, Eileen O'Brien, et al. Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecology Letters*, 7(12):1121–1134, 2004.

Tad Dallas, Andrew W Park, and John M Drake. Predicting cryptic links in host-parasite networks. *PLoS Computational Biology*, 13(5): e1005557, 2017.

D Matthias Dehling and Daniel B Stouffer. Bringing the eltonian niche into functional diversity. *Oikos*, 127(12):1711–1723, 2018.

D Matthias Dehling, Pedro Jordano, H Martin Schaefer, Katrin Böhning-Gaese, and Matthias Schleuning. Morphology predicts species' functional roles and their degree of specialization in plant–frugivore interactions. *Proceedings of the Royal Society B: Biological Sciences*, 283(1823):20152444, 2016.

Daniel F Doak, James A Estes, Benjamin S Halpern, Ute Jacob, David R Lindberg, James Lovvorn, Daniel H Monson, M Timothy Tinker, Terrie M Williams, J Timothy Wootton, et al. Understanding and predicting ecological dynamics: are major surprises inevitable. *Ecology*, 89(4):952–961, 2008.

Jennifer A Dunne, Richard J Williams, and Neo D Martinez. Food-web structure and network theory: the role of connectance and size. *Proceedings of the National Academy of Sciences*, 99(20):12917–12922, 2002a.

Jennifer A Dunne, Richard J Williams, and Neo D Martinez. Network structure and biodiversity loss in food webs: robustness increases

with connectance. *Ecology Letters*, 5(4):558–567, 2002b.

Anna Eklöf, Ute Jacob, Jason Kopp, Jordi Bosch, Rocío Castro-Urgal, Natacha P Chacoff, Bo Dalsgaard, Claudio Sassi, Mauro Galetti, Paulo R Guimarães, et al. The dimensionality of ecological networks. *Ecology Letters*, 16(5):577–583, 2013.

Brent C Emerson and Rosemary G Gillespie. Phylogenetic analysis of community assembly and structure over space and time. *Trends in Ecology and Evolution*, 23(11):619–630, 2008.

Mark C Emmerson and Dave Raffaelli. Predator–prey body size, interaction strength and the stability of a real food web. *Journal of Animal Ecology*, 73(3):399–409, 2004.

Gregory S Gilbert and Campbell O Webb. Phylogenetic signal in plant pathogen–host range. *Proceedings of the National Academy of Sciences*, 104(12):4979–4983, 2007.

Lloyd Goldwasser and Jonathan Roughgarden. Sampling effects and the estimation of food-web properties. *Ecology*, 78(1):41–54, 1997.

José M Gómez, Miguel Verdú, and Francisco Perfectti. Ecological interactions are evolutionarily conserved across the entire tree of life. *Nature*, 465(7300):918–921, 2010.

Nicholas J Gotelli. Null model analysis of species co-occurrence patterns. *Ecology*, 81(9):2606–2621, 2000.

Dominique Gravel, Elsa Canard, Frédéric Guichard, and Nicolas Mouquet. Persistence increases with diversity and connectance in trophic metacommunities. *PLoS ONE*, 6(5):e19374, 2011.

Dominique Gravel, Timothée Poisot, Camille Albouy, Laure Velez, and David Mouillot. Inferring food web structure from predator–prey body size relationships. *Methods in Ecology and Evolution*, 4 (11):1083–1090, 2013.

Roger Guimerà and Marta Sales-Pardo. Missing and spurious interactions and the reconstruction of complex networks. *Proceedings of the National Academy of Sciences*, 106(52):22073–22078, 2009.

Mark O Hill. Diversity and evenness: a unifying notation and its consequences. *Ecology*, 54(2):427–432, 1973.

Matthew C Hutchinson, Edgar Fernando Cagua, and Daniel B Stouffer. Cophylogenetic signal is detectable in pollination interactions

across ecological scales. *Ecology*, 98(10):2640–2652, 2017.

Pedro Jordano. Sampling networks of ecological interactions. *Functional Ecology*, 30(12):1883–1893, 2016.

Bruce Kogut. The stability of joint ventures: Reciprocity and competitive rivalry. *Cooperative Forms of Transnational Corporation Activity*, 13:267, 1994.

GD Kokkoris, AY Troumbis, and JH Lawton. Patterns of species interaction strength in assembled theoretical competition communities. *Ecology Letters*, 2(2):70–74, 1999.

Boris R Krasnov, Miguel A Fortuna, David Mouillot, Irina S Khokhlova, Georgy I Shenbrot, and Robert Poulin. Phylogenetic signal in module composition and species connectivity in compartmentalized host-parasite networks. *The American Naturalist*, 179(4): 501–511, 2012.

Anne E Magurran and Brian J McGill. *Biological diversity: frontiers in measurement and assessment*. Oxford University Press, 2011.

Kevin McCann, Alan Hastings, and Gary R Huxel. Weak trophic interactions and the balance of nature. *Nature*, 395(6704):794, 1998.

Russell E Naisbit, Rudolf P Rohr, Axel G Rossberg, Patrik Kehrli, and Louis-Félix Bersier. Phylogeny versus body size as determinants of food web structure. *Proceedings of the Royal Society B: Biological Sciences*, 279(1741):3291–3297, 2012.

Jens M Olesen, Jordi Bascompte, Heidi Elberling, and Pedro Jordano. Temporal dynamics in a pollination network. *Ecology*, 89(6):1573–1582, 2008.

Kevin J Olival, Parviez R Hosseini, Carlos Zambrana-Torrelio, Noam Ross, Tiffany L Bogich, and Peter Daszak. Host and viral traits predict zoonotic spillover from mammals. *Nature*, 546(7660):646, 2017.

Bruce D Patterson and Wirt Atmar. Nested subsets and the structure of insular mammalian faunas and archipelagos. *Biological Journal of the Linnean Society*, 28(1-2):65–82, 1986.

William D Pearse, Marc W Cadotte, Jeannine Cavender-Bares, Anthony R Ives, Caroline M Tucker, Steve C Walker, and Matthew R Helmus. pez: phylogenetics for the environmental sciences. *Bioinformatics*, 31(17):2888–2890, 2015.

Owen L Petchey and Kevin J Gaston. Functional diversity (fd), species richness and community composition. *Ecology Letters*, 5(3):402–411, 2002.

Owen L Petchey, Andrew P Beckerman, Jens O Riede, and Philip H Warren. Size, foraging, and food web structure. *Proceedings of the National Academy of Sciences*, 105(11):4191–4196, 2008.

Owen L Petchey, Ulrich Brose, and Björn C Rall. Predicting the effects of temperature on food web connectance. *Philosophical Transactions of the Royal Society B*, 365(1549):2081–2091, 2010.

Lauren C Ponisio, Marilia P Gaiarsa, and Claire Kremen. Opportunistic attachment assembles plant–pollinator networks. *Ecology Letters*, 20(10):1261–1272, 2017.

Enrico L Rezende, Pedro Jordano, and Jordi Bascompte. Effects of phenotypic complementarity and phylogeny on the nested structure of mutualistic networks. *Oikos*, 116(11):1919–1929, 2007.

Enrico L Rezende, Eva M Albert, Miguel A Fortuna, and Jordi Bascompte. Compartments in a marine food web associated with phylogeny, body mass, and habitat structure. *Ecology Letters*, 12(8): 779–788, 2009.

Jens O Riede, Björn C Rall, Carolin Banasek-Richter, Sergio A Navarrete, Evie A Wieters, Mark C Emmerson, Ute Jacob, and Ulrich Brose. Scaling of food-web properties with diversity and complexity across ecosystems. *Advances in Ecological Research*, 42:139–170, 2010.

Serguei Saavedra, Daniel B Stouffer, Brian Uzzi, and Jordi Bascompte. Strong contributors to network persistence are the most vulnerable to extinction. *Nature*, 478(7368):233–235, 2011.

Osvaldo E Sala, F Stuart Chapin, Juan J Armesto, Eric Berlow, Janine Bloomfield, Rodolfo Dirzo, Elisabeth Huber-Sanwald, Laura F Huenneke, Robert B Jackson, Ann Kinzig, et al. Global biodiversity scenarios for the year 2100. *Science*, 287(5459):1770–1774, 2000.

Herbert A Simon. On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440, 1955.

Martina Stang, Peter GL Klinkhamer, Nickolas M Waser, Ingo Stang, and Eddy van der Meijden. Size-specific interaction patterns and size matching in a plant–pollinator interaction web. *Annals of Botany*, 103(9):1459–1469, 2009.

Daniel B Stouffer and Jordi Bascompte. Understanding food-web persistence from local to global scales. *Ecology Letters*, 13(2):154–161, 2010.

Daniel B Stouffer, Juan Camacho, and Luís A Nunes Amaral. A robust measure of food web intervality. *Proceedings of the National Academy of Sciences*, 103(50):19015–19020, 2006.

Daniel B Stouffer, Juan Camacho, Wenxin Jiang, and Luís A Nunes Amaral. Evidence for the existence of a robust pattern of prey selection in food webs. *Proceedings of the Royal Society B: Biological Sciences*, 274(1621):1931–1940, 2007.

Daniel B Stouffer, Marta Sales-Pardo, M Irmak Sirer, and Jordi Bascompte. Evolutionary conservation of species' roles in food webs. *Science*, 335(6075):1489–1492, 2012.

Ross M Thompson and Colin R Townsend. Energy availability, spatial heterogeneity and ecosystem size predict food-web structure in streams. *Oikos*, 108(1):137–148, 2005.

David Tilman. The ecological consequences of changes in biodiversity: a search for general principles. *Ecology*, 80(5):1455–1474, 1999.

Jason M Tylianakis, Raphael K Didham, Jordi Bascompte, and David A Wardle. Global change and species interactions in terrestrial ecosystems. *Ecology Letters*, 11(12):1351–1363, 2008.

Brian Uzzi. Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative Science Quarterly*, pages 35–67, 1997.

David I Warton, F Guillaume Blanchet, Robert B O'Hara, Otso Ovaskainen, Sara Taskinen, Steven C Walker, and Francis KC Hui. So many variables: joint modeling in community ecology. *Trends in Ecology and Evolution*, 30(12):766–779, 2015.

John J Wiens and Catherine H Graham. Niche conservatism: integrating evolution, ecology, and conservation biology. *Annual Review of Ecology, Evolution, and Systematics*, 36:519–539, 2005.

CO-AUTHORSHIP STATEMENTS

The following pages contain the co-authorship statements for each chapter of this thesis.

Deputy Vice-Chancellor's Office
Postgraduate Research Office

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:

The entirety of the first chapter was published in the Journal of the Royal Society Interface (doi:10.1098/rsif.2018.0747)

Please detail the nature and extent (%) of contribution by the candidate:

The candidate contributed to the design of the study (60%), performed the analyses (100%), and wrote the first draft (100%). All authors contributed to the subsequent drafts.

Certification by Co-authors:

If there is more than one co-author then a single co-author can sign on behalf of all

The undersigned certifies that:

- The above statement correctly reflects the nature and extent of the PhD candidate's contribution to this co-authored work
- In cases where the candidate was the lead author of the co-authored work he or she wrote the text

Name: Daniel B. Stouffer Signature:  Date: 2019-05-29

Deputy Vice-Chancellor's Office
Postgraduate Research Office

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:

The entirety of the second chapter was published in Nature Communications (doi:10.1038/s41467-018-05056-0)

Please detail the nature and extent (%) of contribution by the candidate:

The candidate contributed to the design of the study (70%), performed the analyses (100%), and wrote the first draft (100%). All authors contributed to the subsequent drafts.

Certification by Co-authors:

If there is more than one co-author then a single co-author can sign on behalf of all

The undersigned certifies that:

- The above statement correctly reflects the nature and extent of the PhD candidate's contribution to this co-authored work
- In cases where the candidate was the lead author of the co-authored work he or she wrote the text

Name: Daniel B. Stouffer Signature:  Date: 2019-05-29

Deputy Vice-Chancellor's Office
Postgraduate Research Office

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:

The entirety of the third chapter was submitted to PLOS Computational Biology (manuscript number PCOMPBIOL-D-18-01356). The manuscript was rejected but the editors allowed us to resubmit. We are currently working on a new version of the manuscript.

Please detail the nature and extent (%) of contribution by the candidate:

The candidate contributed to the design of the study (70%), contributed to the development of the software (50%), performed the analyses (100%), and wrote the first draft (100%). All authors contributed to the subsequent drafts.

Certification by Co-authors:

If there is more than one co-author then a single co-author can sign on behalf of all

The undersigned certifies that:

- The above statement correctly reflects the nature and extent of the PhD candidate's contribution to this co-authored work
- In cases where the candidate was the lead author of the co-authored work he or she wrote the text

Name: Daniel B. Stouffer Signature:  Date: 2019-05-29

Deputy Vice-Chancellor's Office
Postgraduate Research Office

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:

The entirety of the fourth chapter was prepared for submission to Ecology Letters. At the time of the submission of the thesis, the manuscript has not yet been submitted.

Please detail the nature and extent (%) of contribution by the candidate:

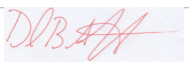
The candidate contributed to the design of the study (80%), performed the analyses (100%), and wrote the first draft (100%). All authors contributed to the subsequent drafts.

Certification by Co-authors:

If there is more than one co-author then a single co-author can sign on behalf of all

The undersigned certifies that:

- The above statement correctly reflects the nature and extent of the PhD candidate's contribution to this co-authored work
- In cases where the candidate was the lead author of the co-authored work he or she wrote the text

Name: Daniel B. Stouffer Signature:  Date: 2019-05-29

Deputy Vice-Chancellor's Office
Postgraduate Research Office

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:

The entirety of the first appendix was published in Functional Ecology (doi:10.1111/1365-2435.13237)

Please detail the nature and extent (%) of contribution by the candidate:

Matthew C. Hutchinson led the writing of the manuscript, and the candidate contributed to the design of the study (40%) as well as the writing (20%).

Certification by Co-authors:

If there is more than one co-author then a single co-author can sign on behalf of all

The undersigned certifies that:

- The above statement correctly reflects the nature and extent of the PhD candidate's contribution to this co-authored work
- In cases where the candidate was the lead author of the co-authored work he or she wrote the text

Name: Daniel B. Stouffer Signature:  Date: 2019-05-29

Part VI

APPENDIX

A

SEEING THE FOREST FOR THE TREES: PUTTING MULTILAYER NETWORKS TO WORK FOR COMMUNITY ECOLOGY

Matthew C. Hutchinson[1], Bernat Bramon Mora[2], Shai Pilosof[3], Allison K. Barner[4], Sonia Kéfi[5], Elisa Thébault[6], Pedro Jordano[7], Daniel B. Stouffer[2]

1 Department of Ecology and Evolutionary Biology, Princeton University, Princeton, Princeton NJ, USA.

2 Centre for Integrative Ecology, School of Biological Sciences, University of Canterbury, Christchurch, New Zealand.

3 Department of Ecology & Evolution, University of Chicago, Chicago, IL, USA.

4 Department of Environmental Science, Policy, and Management, University of California Berkeley, Berkeley CA, USA.

5 Institute Of Science De L'évolution De Montpellier, University of Montpellier, Montpellier, France.

6 Institute of Ecology and Environmental Sciences of Paris, Sorbonne Université, Paris, France

7 Department of Integrative Ecology, Estación Biológica de Doñana, Sevilla, Spain

A.1 Abstract

1. A framework for the description and analysis of multilayer networks is established in statistical physics and calls are increasing for their adoption by community ecologists. Multilayer networks in community ecology will allow space, time, and multiple interaction types to be incorporated into species-interaction networks.

2. While the multilayer-network framework is applicable to ecological questions, it is one thing to be able to describe ecological communities as multilayer networks and another for multilayer networks to actually prove useful for answering ecological questions. Importantly, documenting multilayer network structure requires substantially greater empirical investment than standard ecological networks. In response, we argue that this additional effort is worthwhile and describe a series of research lines

where we expect multilayer networks will generate the greatest impact.

3. Inter-layer edges are the key component that differentiate multilayer networks from standard ecological networks. Inter-layer edges join different networks—termed layers—together and represent ecological processes central to the species interactions studied (e.g. inter-layer edges representing movement for networks separated in space). Inter-layer edges may take a variety of forms, be species- or network-specific, and be measured with a large suite of empirical techniques. Additionally, the sheer size of ecological multilayer networks also requires some changes to empirical data collection around interaction quantification, collaborative efforts, and collation in public databases.

4. Network ecology has already touched on a wide swathe of ecology and evolutionary biology. Because network stability and patterns of species linkage are the most developed areas of network ecology, they are a natural starting place for multilayer investigations. However, multilayer networks will also provide novel insights to niche partitioning, the connection between traits and species' interactions, and even the geographic mosaic of coevolution.

5. *Synthesis:* Multilayer networks provide a formal way to bring together the study of species-interaction networks and the processes that influence them. However, describing inter-layer edges and the increasing amounts of data required represent challenges. The payoff for added investment will be ecological networks that describe the composition and capture the dynamics of ecological communities more completely and, consequently, have greater power for understanding the patterns and processes that underpin diversity in ecological communities.

A.2 Introduction

It has been stated that "no theory of the cell, of social media or of the Internet can ignore the profound network effects that their interconnectedness cause" (Barabási, 2011). Many ecologists would argue that one paradigmatic example is noticeably missing from this list: the ecological community. After all, interactions are a defining characteristic of every species' "millieu" since no individual organism exists without participating in some sort of ecologically relevant interspecific interaction during its lifetime. Interactions between species underpin community ecology to the extent that they have even been claimed to

represent "the architecture of biodiversity" (Bascompte and Jordano, 2013).

Indeed, scholarly accounts of the myriad connections between species date back at least to al-Jāhiz in the 9th century or Aristotle in the 4th century BCE (Egerton, 2012). More pertinently, network ecology has seen massive growth in the past 40 years. Fundamental food-web theory has been built on with empirical observation and experimentation (Pascual et al., 2006), a vibrant literature on bipartite interaction networks has emerged around mutualisms (Bascompte and Jordano, 2013), and models to predict species linkage have gained realism and accuracy (Eklöf et al., 2013). While this literature has generated steady progress in our understanding of ecological communities, it has relied on data representing single points in space and time, or a narrow picture of a guild's biotic interactions. Collecting sets of networks has been one way to overcome this issue (Ponisio et al., 2017) but does not account for the dependency between networks. A new tool—multilayer networks—has recently entered the fray with the promise of addressing these issues but key questions around their use remain unresolved.

The theoretical framework for multilayer networks (Boccaletti et al., 2014; Kivelä et al., 2014) and their implementation in ecology (Pilosof et al., 2017) already exist. Indeed, theoretical contributions are already helping to bridge multilayer networks and ecological theory (García-Callejas et al., 2018; Godoy et al., 2018). Briefly, multilayer networks provide a way to encode different entities and/or interactions in a single network by distinguishing between types of network edges (Boccaletti et al., 2014; Kivelä et al., 2014). Ecological multilayer networks (hereafter, EMNs) will typically represent species interactions through time and space, or across types of interactions (Pilosof et al., 2017). For example, spatial and temporal EMNs can consist of several networks of a focal species-interaction type (e.g. plant-pollinator or host-parasite interactions; with each network referred to as a layer composed of intra-layer edges) linked to each other via inter-layer edges that describe an additional ecological process, such as gene flow between habitats or change in abundance through time. EMNs composed of different types of species interactions (e.g. pollination, seed-dispersal, folivory) can also be represented as multilayer networks (hereafter, multiplex EMNs). Inter-layer edges in multiplex EMNs may represent the effect of one interaction type on the other (but see Gracia-Lázaro et al. 2018). Using both intra- and inter-layer edges, EMNs enable multiple descriptors of ecological communities to be linked together into a single network. That network will capture both interactions between species as well as a factor (e.g. abundance, movement, physiology) helping to determine those interactions.

Ecology has always championed the notion that complexity should never be embraced for complexity's sake alone, except when the tools to face the challenge are available (Green et al., 2005) and the alternative is an unpalatable sacrifice of biological realism (May, 1972). With this in mind, the EMN framework represents an important step towards capturing the true complexity of ecological communities. The methodological tools to analyse these networks are already established (Boccaletti et al., 2014; Kivelä et al., 2014); as a technical advance, the power of multilayer networks is therefore clear. However, a methodological framework to describe and study EMNs is not enough to satisfy the all important question; 'so what?'. While all methodological innovations require use to be broken in and prove their value, it is still unclear how describing EMNs will improve our understanding of ecological communities. To guide this process, we identify several considerations for the empirical implementation of EMNs and outline a series of research areas that we believe will most immediately benefit.

A.3 Inter-layer edges: why, when, how

Inter-layer edges connect nodes between layers (usually these are between a species and its representation in each layer but see Gracia-Lázaro et al. 2018) and they are the key component that differentiates multilayer networks from previous approaches to ecological networks. Incorporating inter-layer edges into ecological networks allows ecological processes that influence interaction patterns to be incorporated into network structure creating a more complete description of community dynamics. These edges may represent a myriad of processes, and their definition will depend on the specifics of the system and questions at hand. Because inter-layer edges are essential to a multilayer approach, we discuss three main considerations for their use: (i) Do inter-layer edges add necessary information? (ii) Can inter-layer edges have the same weight across the network or should they be specific to pairs of nodes? (iii) How can inter-layer edges be quantified?

A.3.1 Why use inter-layer edges and EMNs?

In many cases, inter-layer edges will represent ecological processes familiar to community ecologists. For example, in spatial EMNs (Fig. 16d), inter-layer edges can represent movement between layers, similar to meta-communities; in temporal EMNs (e.g. Costa et al. 2018; Fig. 16c), they can represent changes in abundance, echoing the interplay between abundance and foraging inherent in population and functional response theories; in pathogen systems, they can represent temporal genetic changes, similar to phylogenetic trees in phylody-

namics (Pilosof et al., 2018). That each of the processes represented by inter-layer edges already have theoretical frameworks built around them begs the question: why study them with an EMN approach?

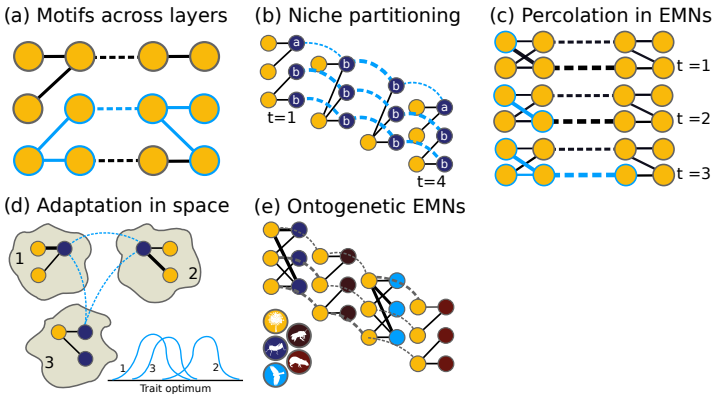


Figure 16: Examples where we expect EMNs to be innovative for community ecology. (a) How interaction patterns within layers are connected between them can give definition to the interconnectedness of networks across interaction types. Here, 3-species motifs are highlighted (blue outline) as one pattern worth investigating. (b) Identifying groups of species across an EMN provides evidence for the consistency of niche partitioning. Here, letters represent hypothetical groups based on inter- and intra-layer edges. (c) Percolation across EMNs can be used to study how ecosystems will respond to perturbation. Blue outlining marks the trajectory of percolation here and a key question is how inter-layer edges representing ecological processes will transmit perturbations. (d) Three habitat patches are represented and populations of one species (upper dark blue circle) are connected between them. Different interaction patterns between layers determine functional-trait optima and inter-layer edges may be key to adaptation across the system. (e) A multiplex network of seasonal attachment by a tree population to different animal guilds is represented. Interaction patterns within and across layers could be used to determine individual fitness outcomes.

First, we already know that, for example, dispersal between sites can maintain local resource populations—and therefore consumer-resource interactions—in areas where they would otherwise become locally extinct (Rosenheim, 2001), or that changes in resource species' density alters consumer behaviour (Holling, 1959). EMNs provide a way to integrate existing knowledge of this sort into the structure of ecological networks for a more holistic study of species interactions. Second, by combining species interactions (intra-layer edges) with existing knowledge, as inter-layer edges, the types of questions that will be addressed with ecological networks become more salient. With a set of networks distributed in space, the logical question might be to understand how these networks differ based on their proximity. Instead if those networks are studied as a single spatial EMN, for which proximity-based effects are structurally explicit, then the questions that are most intuitive to ask become those that pertain

to population and community dynamics, biogeography, and ecosystem management (Figs. 16 & 17). A sharper focus on these areas can improve the ability of network ecology to speak to fundamental questions around biodiversity maintenance and their applied counterparts in conservation.

A.3.2 When do inter-layer edges need to be quantified?

Inter-layer edges may take several forms from none and uniform edge weights (including binary weights) to non-uniform edges. EMNs with no inter-layer edges are a set of independent networks (identical to previous approaches in network ecology that have tried to assess spatial or temporal variation; e.g. Poisot et al. 2012; Cirtwill et al. 2018). When EMNs have uniform inter-layer edges (i.e. all edges between layer have the same weight), they assume that all connections between all layers have the same effect and inter-layer edges effectively describe the assumed dependence of layers on each other. Finally, if inter-layer edges are non-uniform (i.e. each link between nodes in different layers will be distinct), they represent the actual or estimated linkage between two nodes and add substantial resolution to the network. EMNs with different types of inter-layer edges can be used to ask different questions (Table 1) even though the most complete description of the community will require non-uniform inter-layer edges.

When inter-layer edges have been used, they are often uniform (Kéfi et al., 2016; Timóteo et al., 2018). The strength of uniform inter-layer edges is an assumption about the interdependence of layers. If inter-layer edge strength is much greater than intra-layer strengths, inter-layer processes dominate the community and determine structural properties; and vice versa (Pilosof et al., 2017). In Gorongosa National Park, seed-dispersal modules become habitat-specific when the inter-layer weighting is small relative to intra-layer weights, however when inter-layer weightings were large, landscape-level groupings of plant-disperser interactions emerged (Timóteo et al., 2018). These results naturally suggest that when layers (in this case, habitats) are considered to be distinct ecological units, species interactions become grouped by those, often arbitrary, borders. However, when a more continuous view is taken—as is done in Timoteo et al. 2018 when inter-layer edges become large relative to intra-layer edges—a holistic image of those interactions across the landscape, community, or timespan emerges.

Shifting from uniform to non-uniform inter-layer edges in EMNs can provide more detailed information on the processes that operate between layers and their role in community structure. In a spatial EMN in which inter-layer edges represent inter-patch relationships

Box 1: Multilayer networks for ecosystem management

The impetus for an EMN approach to ecology stems from the idea that they can capture a greater portion of ecosystem complexity (Fontaine et al., 2011; Pilosof et al., 2017), and capturing such complexity is particularly necessary for ecosystem management. EMNs offer several opportunities for applied ecology. More than a decade ago, the Millennium Ecosystem Assessment identified four key anthropogenic drivers of global change: climate change, nutrient pollution, invasive species, and land-use change (Millennium Ecosystem Assessment, 2005). Here, we highlight potential uses of an EMN approach to ecosystem management by exploring examples pertinent to each of these four drivers.

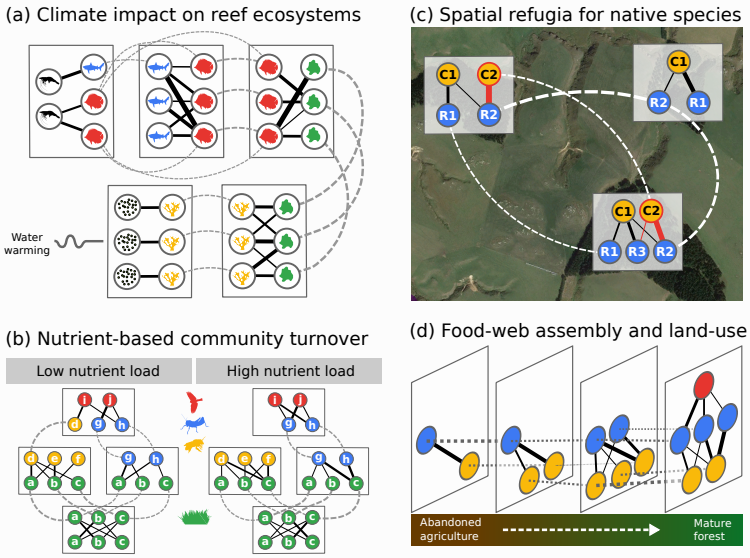


Figure 17: EMNs and global change ecology. (a) A multiplex EMN of a coral reef. Intra-layer edges describe ecological interactions such as competition between coral and algae. Inter-layer edges represent the per species fitness effects of one interaction type on the other. (b) Two hypothetical multiplex EMNs for low- and high-nutrient loads show how changes in nutrient load may cascade upwards from plant associations to higher webs (colours represent guilds; red: insectivores, yellow: pollinators, blue: herbivores, green: plants). (c) Food webs for three habitat fragments with dispersal as inter-layer edges (white dotted lines). In this example, C2 is invasive and specialises on R2. Inter-layer edges between patches allows R2 to persist. Map data: Google, DigitalGlobe. (d) Food web assembly as land-use changes from agriculture to forest. Circles are different species and colour indicates their trophic level. Inter-layer edges depict change in abundance between time points.

The warming of the world's oceans is of particular concern for coral reefs and the ecological associations (Sale, 2013) that coral-zooxanthellae symbiosis supports. A multiplex EMN—where layers are interaction types such as coral-zooxanthellae association, coral-algae competition, herbivory, and cleaning mutualisms while inter-layer edges represent the dependence of interaction types on each other—could be used to assess how damage to coral-zooxanthellae associations might cascade to the rest of the community (Fig. 17a). Nutrient pollution can alter plant communities (Hautier et al., 2009) and potentially the pollinator, herbivore, and predator-prey webs they support. Multiplex EMNs founded upon plant associational interactions could be used to examine how nutrient-driven changes in plant community composition will predict ecosystem service provisioning by animals based on percolation from the plant-plant layer to higher trophic layers (Fig. 17b). The proliferation of non-native species is reality in a globalised world and establishment is influenced by trophic interactions (Maron and Vilà, 2001; Salo et al., 2007). An invasion biology application of EMNs might take a meta-community approach and consist of habitat-fragment layers, inter-layer edges encoding species-specific dispersal between fragments, and be used to forecast native resistance and define management actions (Fig. 17c). Land conversion for anthropogenic purposes is rampant (Ramankutty et al., 2008) and can profoundly impact species' abundance and ecological interactions (Ponisio et al., 2017). A temporal EMN would allow the dynamics of community decomposition or assembly under land-use change to be precisely identified. Each layer might be a food web at a certain point in time with inter-layer edges representing change in abundance and will be useful for determining how ecological community outcomes are linked to land-use change (Fig. 17d).

(e.g. species-specific movement or predation risk differences between patches; Fig. 16d), non-uniform inter-layer edges will explicitly incorporate variation in inter-patch movement between species and populations. This variation is crucial because it will eventually affect within-patch population dynamics (Rosenheim, 2001). Likewise in temporal networks constructed for highly seasonal ecosystems, inter-layer edges representing fluctuations in species' abundance are bound to be non-uniform because uniform inter-layer edges will ignore the species-specific responses to seasonality and therefore distort conclusions about how temporal resource competition or availability structures the community (Costa et al., 2018). Accordingly, if inter-layer edges linking plant species across seasons are uniform they assume that the effect of season on plant biomass is equivalent across species despite the fact that annuals may switch from low to high biomass while a woody perennial maintains biomass between seasons (e.g. Singh and Yadava 1974).

A.3.3 How can inter-layer edges be quantified?

Measuring inter-layer edges will often be straightforward given ecology's extensive empirical toolbox. Community sampling tools such as camera-trapping grids (and other community sampling methods) can be particularly useful for estimating species' relative abundance given their ability to yield data for large numbers of species with relatively little time investment. Recent developments (Norouzzadeh et al., 2018) mean that extracting occupancy measures from camera-trap data represents a low cost and low intensity way to estimate inter-layer edges for both spatial and temporal EMNs. Similarly, GPS tags, population-genetic monitoring, and mark-recapture techniques can estimate species-wise connectivity between layers in spatial EMNs (Table 1). In temporal EMNs, where time points are modelled as layers, species' abundance and biomass are modelled as inter-layer edges and can be estimated with a variety of standard approaches (Table 1).

Non-uniform inter-layer edges in multiplex EMNs are less obvious but an intuitive representation would be the effect of one interaction type on another. These might be estimated by applying specific natural history knowledge of the species that connect layers. For example, a multiplex EMN where one layer represents predator-prey interactions and another represents plant-herbivore interactions could have inter-layer edges that describe how the landscape of fear alters prey/herbivore habitat-use. A landscape-of-fear-derived measure would capture the way that predation on herbivores alters their foraging patterns therefore linking herbivory patterns to predation interactions. Similarly, in the case of dispersers and predators of seeds, inter-layer edges representing interference between these two guilds

might be measured as the efficiency of fruit removal by dispersers, which would deny foraging opportunities for seed-predators locally.

If empirical quantification is impossible, inter-layer edges may also be inferred by combining natural history and ecological theory. Morphological traits are intricately linked to ecological network structure and this is particularly true for body size (Brose et al., 2006). In addition, larger bodied species tend to have larger home ranges and dispersal ability (Lindstedt et al., 1986; Bowman et al., 2002). Therefore, in spatial EMNs, combining body size with distance between locations (layers) could be used to generate species- and layer-specific dispersal potentials.

It is also possible to infer inter-layer edges based on intra-layer edges. Neighborhood flow coupling—where inter-layer weights are designated based on the similarity of a node's intra-layer edges between two layers—provides one approach to do so (Aslak et al., 2018) and would be appropriate for temporal and spatial EMNs. For example, if a plant is visited by the same pollinators in two consecutive years, this approach would generate a strong inter-layer link between the plant species in those two layers likely implying maintenance of abundance and pollination service between years. A similar approach could infer inter-layer edges representing the change in abundance of species i between a layer at time t and a layer at time $t+1$ based on the net effect of i's intra-layer edges at t. Inter-layer edges with inferred weights should not remove the incentive to measure inter-layer dependencies but may provide more information to an analysis than uniform inter-layer edges because, even if they only capture coarse differences between species (e.g. evergreen vs. deciduous), they will bring greater natural history into the structure of EMNs.

A.3.4 Additional considerations when describing, quantifying, and collating EMNs

The quantification of species interactions is a much-discussed facet of community ecology (Novak et al., 2016). In an EMN scenario, where different interaction types are represented, an additional challenge is to quantify interactions in a common currency and has thus far proven difficult (García-Callejas et al., 2018). The per capita effect of one species on the other is key to theory that underlies food webs (May, 1972) and becomes the gold standard since it is not specific to an interaction type (Novak et al., 2016). Several approaches exist for quantifying interactions in this way and these need not always be experimental (Bimler et al., 2018; Lanuza et al., 2018). Given the importance of per capita effects in ecological theory, it is important that these emerging approaches are built on to help generate accurate species-interaction data because, with increasing richness, exper-

Table 1: Quantifying inter-layer edges in multilayer networks

Inter-layer edge	Network Type	Significance	Quantifying Edges
Dispersal between patches	Spatial EMN	Movement of individuals between patches creates a correlation in the structure of those layers. Inter-layer edges weighted by dispersal recognise movement as a determinant of interspecific interaction patterns.	GPS tracking of individuals, camera trapping in dispersal corridors, population genetic monitoring of patches, or mark-recapture.
Difference in abundance	Spatial and temporal EMNs	Abundance influences interspecific interactions. Inter-layer edges representing abundance allow the EMN to capture population and community trends simultaneously.	Community monitoring within layers. For animals, trapping grids, eDNA monitoring, point counts, or mark-recapture. For plants, quadrat- or transect-based surveys of biomass and eDNA monitoring.
Dependency of species interaction types	Multiplex EMN	Species' participation in distinct interaction types and their effect on each other determine network structure.	Experimental determination of fitness differences in the presence of one or both interactions. Proxies for fitness differences are more realistic.
Change in energetic requirements	Individual-based temporal EMN	Energetic requirements can change through time. Pregnancy and lactation can double an animal's energetic requirements and alter foraging decisions.	Expected change in energetic requirements. May be quantified through a combination of individual physiological status and literature-derived values for the effect of changing state.
Changing predation/parasitism risk.	Spatial and temporal EMNs	Risk can change through space and time. Inter-layer edges capturing differences in risk between layers can link direct and trait-mediated consumer effects.	Environmental monitoring of parasite transmission risk and predator populations.
Spatial subsidies	Spatial EMN	Ecosystems and communities do not exist in isolation and inter-layer edges can represent the flow of nutrients from one patch to the other.	Environmental monitoring between patches.

imental approaches quickly become impractical, however, there is still substantial work to be done (Barner et al., 2018).

Alternatively, interaction frequency has been used as a proxy for per capita interaction strength (Vázquez et al., 2005). However, it is unclear how comparable the frequencies of two interaction types are to each other. For example, it is unclear how a bird eating five fruits is comparable to five parasites being present in that bird's gastrointestinal tract. While transforming raw frequencies to relative frequencies would be one way to make these numbers more comparable, a potential common currency for species interactions—relative read abundance (RRA)—goes a step further. RRA quantifies interactions from DNA meta-barcoding studies and is the proportion of a sample, perhaps pollen on a bee's legs or plant material in an ungulate's dung, assigned to each resource taxon (Kartzinel et al., 2015). Because read abundance is a count of DNA barcodes belonging to each taxon, RRA counts ecological interactions in the same way regardless of interaction type. DNA meta-barcoding is particularly relevant to trophic interactions and even difficult-to-describe interactions such as intra-guild predation can feasibly be studied (Toju and Baba, 2018). Non-trophic interactions require additional study to identify how they might be estimated with DNA barcoding. For example, belowground plant species' RRA (Hiiesalu et al., 2012) should be examined in relation to plant-plant competition. The wide-ranging applications of DNA meta-barcoding to species interactions, declining cost, and consistent approach to quantifying interactions make RRA a potential solution to empirically constructing the intra-layer edges of multiplex EMNs in a standardised way.

Even when harnessing technological advances to estimate inter- and intra-layer edges, substantial boots-on-the-ground effort is required to collect EMNs. Fortunately, several solutions to the problem of big data collection exist. Foremost, large-scale collaborations represent an ideal facilitator of such endeavours. Ecological questions of comparable complexity have been embraced through massive collaborative efforts such as the Nutrient Network experiment and macroecological working groups. It is also important that EMNs are collated in open-access databases since ecological network databases have been crucial to the identification of patterns in ecological networks (e.g. Olesen et al. 2007; Hutchinson et al. 2017). The existence of platforms that curate and standardise data such as mangal (http://mangal.io/; Poisot et al. 2016) and globi (https://www.globalbioticinteractions.org/; Poelen et al. 2014) will streamline the collation and curation of EMNs as well as facilitate their use in identifying patterns in community organisation.

A.4 Putting EMNs to work in community ecology

Now that we have described the considerations involved in obtaining EMN data, we turn to several areas of active research in community ecology where the application of EMNs can lead to conceptual advances. For each area below, we briefly describe a few relevant questions and hypotheses an EMN approach can address and the ways in which it may be implemented.

A.4.1 The structure of species interaction networks

The search for non-random structural patterns in ecological networks has generated some of the clearest evidence for consistency in the organisation of ecological communities. Insights into niche partitioning (Beckerman et al., 2006), forbidden interactions (Olesen et al., 2010), and sampling of interactions (Jordano, 2016) have been garnered from the observation that rarely more than 30-40% of possible interactions are realised in food webs (Fortuna et al., 2010). In EMNs, exploring basic structural properties and, in particular, the linkage patterns between inter-layer edges and intra-layer structure deserves attention (Table 1). One area of particular interest with respect to network structural patterns is the way that plants are connected to the multitude of primary consumer guilds that they support. Initial findings suggest that plants do connect herbivores and pollinators in non-random ways (Sauve et al., 2016) and this should be expanded to other primary consumer groups too.

The identification of structural patterns in EMNs requires null models—which evaluate empirical structure with comparison to random expectation—within which to assess the consistency of observed patterns. In a EMN case, a null model can focus on the placement of inter-layer edges and the ordering of layers as well as intra-layer edges (Bassett et al., 2011). Null models addressing particular EMN components can be particularly useful in isolating the mechanisms producing structural patterns (Bassett et al., 2011). The value of shuffling intra- and inter-layer edges at the same time deserves attention as well as the comparison of shuffling inter-layer edge weights vs. edge presence.

The development of null models for EMNs and the study of overall network metrics, however, should focus on answering specific ecological questions. To this end, there are some computational tools that could provide valuable insights (see Pilosof et al. 2017 for a list of metrics). In particular, the study of network motifs—the suite of n-node interaction patterns within networks—has already proven to be a useful tool in characterising the structural roles that species play in networks (Bramon Mora et al., 2018) and could provide an interest-

ing perspective on how interaction patterns are linked across layers (Fig. 16a). Studying which motifs are linked to which across layers could shed light on the dynamics of the interaction patterns and the consistency of species' roles (as defined by motif membership) across layers. We predict that strong inter-layer edge weights will be associated with species' role consistency across layers because stronger inter-layer edges will tend to represent the relatedness or dependency of the focal species' interactions in each layer.

A.4.2 *Niche partitioning across time and space*

Resource partitioning is central to favoured explanations of diversity in ecological communities (MacArthur and Levins, 1967) and is often observed (McDonald, 2002). Ecological networks have been used to examine niche partitioning for diverse communities (Blüthgen et al., 2008) but many network datasets comprise a small spatial or temporal window. Characterizing resource partitioning within communities in a spatial or temporal EMN will provide an explicit link to models of resource competition and generate a comprehensive picture of specialisation that simultaneously considers variation across space and time as well as the process encoded in the inter-layer edges which will usually be related to population change.

Large mammalian herbivores in Africa are an illustrative example since their resource use has been studied thoroughly. These species tend to be either grazers (grass feeders) or browsers (woody-plant feeders), with few species being mixed feeders (Hempson et al., 2015). However, plant communities in savanna ecosystems are highly seasonal and it is unclear how strictly grazers and browsers separate. A temporal EMN would allow novel access to this question if layers represent mammal-plant herbivory networks at different times of year and inter-layer edges represent changes in population density across the year (Fig. 16b). We expect the group structure of this EMN would be more inclusive in periods where resources are abundant and dissipates to more traditional grazer-browser separation during low resource availability. These results, which describe the behavioural plasticity in large mammal herbivore foraging, would reveal the intensity of resource limitation on such communities and their ability to rewire their diets to cope.

A.4.3 *Network stability*

Central to the adoption of a network approach in ecology was the question of whether diverse systems of interacting species can be stable (May, 1972). In monolayer networks, stability is typically studied as resilience (Thébault and Fontaine, 2010), robustness (Dunne and Williams, 2009), or local stability (Allesina and Pascual, 2008). EMNs

provide a more integrated view of the communities studied and alternate methods must be developed for EMN stability analysis (but see Reis et al. 2014; Danziger et al. 2016). In a recent meta-community study—which studied a spatial EMN in all but name—May's original formulation of local stability was expanded to include an additional matrix of dispersal rates, finding that dispersal can increase stability of meta-ecosystems (Gravel et al., 2016). These results highlight the cross-over between existing theory of ecological meta-communities and EMNs while also emphasising the importance of transitioning to larger-scale approaches to inform ecosystem management decisions. Excitingly, the approach of Gravel et al. (2016) is transferable to other types of multilayer networks if their dispersal matrix is generalised to represent any inter-layer edge.

The spread of disturbances—percolation—through ecological networks has also generated a vibrant literature, particularly in the case of extinction cascades (Dunne and Williams, 2009). Percolation is even more pertinent in multilayer networks since percolation between layers could be particularly relevant to the ecosystem management scenarios that EMNs capture. Several models exist from statistical physics of interdependent networks (Buldyrev et al., 2010; Huang et al., 2011; Liu et al., 2018) and others are needed to understand how percolation proceeds when inter-layer edges represent different processes or are non-uniform (Fig. 16c). Pressingly, different parts of ecosystems can be represented by EMNs that could help identify key edges that transmit perturbations and provide specific management directives. Since modules buffer percolation in monolayers networks (Gaiarsa and Guimaraes Jr, in press), we expect that extinction cascades within layers of, for example, a spatial EMN will remain localised if sink populations are affected, or if source populations are affected extinction will percolate across layers.

A.4.4 Interaction-driven adaptation

Understanding how species interactions drive adaptation in their participants is increasingly important (Galetti et al., 2013). Direct and indirect interactions in ecological networks can have substantial effects on coevolutionary outcomes (Guimarães Jr et al., 2017). An extension of these findings is to explore how population-level adaptation occurs across a multiplex EMN when the species being studied are involved in a variety of interaction types concurrently (Fontaine et al., 2011).

The close matching between plant floral traits and the bill morphology of flower-visiting birds such as hummingbirds (Cotton, 1998) reflects the importance of their association. However, hummingbirds also have predators (Stiles, 1978), parasites (Moens et al., 2016), and competitors (Laverty and Plowright, 1985). Representing different as-

pects of hummingbird ecology as a multiplex EMN allows these interactions and their selective pressures to be studied together. We expect that an adaptive interplay between flower-visitation and competition layers will occur on a multiplex EMN due to strong inter-layer edges (visitation can determine competition and vice versa) that leads to selection on bill morphology and adaptive rewiring of intra-layer edges and their weights in both layers.

A spatial setting is also pertinent to trait adaptation on EMNs. The geographic mosaic of coevolution (Thompson, 2005) describes coevolution occurring across a landscape of patches and a spatial-EMN framework is an ideal tool to study these patches as an integrated whole. Inter-layer edges representing either trait similarity between populations in different layers or gene flow between layers would generate a description of the geographic mosaic that would then allow models of trait adaptation to make predictions about coevolution at a landscape level (Fig. 16d).

A.4.5 Predicting interactions from species traits

Trait-based ecology has made explicit the link between species' morphology and ecological processes (McGill et al., 2006), and has been successfully applied in reducing the dimensionality of complex ecological networks (Eklöf et al., 2013; Laigle et al., 2018). Extrapolating these results, a suite of traits could be leveraged to predict multiplex EMNs. For example, plants face a strong trade-off between competitive ability and herbivory defense (Kempel et al., 2011) and plant secondary compounds influence both herbivory and pollination (Georgelin and Loeuille, 2016). Initial explorations suggest that these correlations may underlie entire communities (Sauve et al., 2016).

While using several traits to predict multiplex EMNs across the web of life is likely within reach, it is also possible that there are individual 'skeleton key' traits that can predict multiplex EMN structure alone. One promising avenue is the nutritional value of resources. Traits representing some measure of nutritional value—be they body size, the leaf economic spectrum, or biochemical composition—have been consistently powerful for community ecology. We expect that the description of a standardised nutritional-quality trait will predict multiplex network structure better than existing models. Such a trait would be a massive coup for community ecology because it would capture a common feature underpinning the organization of ecological communities.

A.4.6 Ontogenetic shifts in species interactions

For many species, the suite of potential ecological interaction partners is not the same throughout an individual's life. Predators will eat bigger prey as they grow, parasites often infect different hosts throughout their lifecycle, and lepidopterans switch from being herbivores to mutualists. Importantly, incorporating ontogeny into food webs yields networks less-robust to extinctions (Rudolf and Lafferty, 2011). In a scenario as distinct as lepidopteran metamorphosis, considering the caterpillar-plant and butterfly-plant associations together as layers of multiplex network could be used to explore their impact on plant communities across the lifecycle of these insects (Astegiano et al., 2017).

Individual-based networks, where the same individual "builds" different ecological interactions with multiple partners through time, are also a natural place for ontogenetic EMNs to be explored. A recent study demonstrated how nurse plant age is positively related to the nestedness of their associational-plant networks (Losapio et al., 2018). Similarly, trees in a population interact sequentially across the growing season with folivores, pollinators, pre-dispersal seed predators, frugivores, post-dispersal seed predators. Multiplex EMNs could be used to study this succession where inter-layer edges represent the proportion of individual investment progressing to the next stage of reproduction (e.g. proportion of flowers pollinated, proportion of seeds dispersed). We predict that the group structure in these multiplex EMNs (Fig. 16e) will be related to individual fitness and may be used to forecast the generational shifts in allele frequencies of the population. Moreover, we expect that intra-layer edge structures will be predictive of inter-layer weights (an individual visited in the greatest frequency and by the greatest diversity of pollinators will contribute a larger investment to the next stage of reproduction).

A.5 Conclusions

Ecological communities and meta-communities are complex systems that harbour immense diversity. Through time, methodological developments in ecology have allowed increasing amounts of that diversity to be described and analysed simultaneously. Ecological networks representing static images of ecological communities have yielded important findings on many aspects of community organisation. Multilayer networks appear to be the logical next step in that expanding toolbox because EMNs make it possible for additional realism—such as population dynamics across space and time as well as multiple types of species interactions—to be built into the structure of species-interaction networks.

Of course, increasing the realism of already data-rich analyses necessitates greater empirical effort to generate interaction networks. Fortunately, the possibilities inherent in emerging data collection technologies, such as DNA meta-barcoding, camera trapping, and remote sensing, combined with tried-and-tested field observations and increasing accessibility of ecological data mean that leveraging EMNs to tackle fundamental questions in community ecology is an achievable ambition. Ultimately, EMNs have the potential to provide the larger-scale and detailed descriptions of ecological communities that may help address historical laments in community ecology (Paine, 1988; Lawton, 1999). Looking forward, it is impossible to determine the impact that EMNs will have on community ecology, but the prospects are bright. All that remains is for boots to hit the ground.

BIBLIOGRAPHY

Stefano Allesina and Mercedes Pascual. Network structure, predator–prey modules, and stability in large food webs. *Theoretical Ecology*, 1(1):55–64, 2008.

Ulf Aslak, Martin Rosvall, and Sune Lehmann. Constrained information flows in temporal networks reveal intermittent communities. *Physical Review E*, 97(6):062312, 2018.

Julia Astegiano, Florian Altermatt, and François Massol. Disentangling the co-structure of multilayer interaction networks: degree distribution and module composition in two-layer bipartite networks. *Scientific Reports*, 7(1):15465, 2017.

Albert-László Barabási. The network takeover. *Nature Physics*, 8(1):14, 2011.

Allison K Barner, Kyle E Coblentz, Sally D Hacker, and Bruce A Menge. Fundamental contradictions among observational and experimental estimates of non-trophic species interactions. *Ecology*, 99(3):557–566, 2018.

Jordi Bascompte and Pedro Jordano. *Mutualistic networks*, volume 70. Princeton University Press, 2013.

Danielle S Bassett, Nicholas F Wymbs, Mason A Porter, Peter J Mucha, Jean M Carlson, and Scott T Grafton. Dynamic reconfiguration of human brain networks during learning. *Proceedings of the National Academy of Sciences*, 108(18):7641–7646, 2011.

Andrew P Beckerman, Owen L Petchey, and Philip H Warren. Foraging biology predicts food web complexity. *Proceedings of the National Academy of Sciences*, 103(37):13745–13749, 2006.

Malyon D Bimler, Daniel B Stouffer, Hao Ran Lai, and Margaret M Mayfield. Accurate predictions of coexistence in natural systems require the inclusion of facilitative interactions and environmental dependency. *Journal of Ecology*, 106(5):1839–1852, 2018.

Nico Blüthgen, Jochen Fründ, Diego P Vázquez, and Florian Menzel. What do interaction network metrics tell us about specialization and biological traits. *Ecology*, 89(12):3387–3399, 2008.

Stefano Boccaletti, Ginestra Bianconi, Regino Criado, Charo I Del Genio, Jesús Gómez-Gardenes, Miguel Romance, Irene Sendina-Nadal, Zhen Wang, and Massimiliano Zanin. The structure and dynamics of multilayer networks. *Physics Reports*, 544(1):1–122, 2014.

Jeff Bowman, Jochen AG Jaeger, and Lenore Fahrig. Dispersal distance of mammals is proportional to home range size. *Ecology*, 83 (7):2049–2055, 2002.

Bernat Bramon Mora, Dominique Gravel, Luis J Gilarranz, Timothée Poisot, and Daniel B Stouffer. Identifying a common backbone of interactions underlying food webs from different ecosystems. *Nature Communications*, 9(1):2603, 2018.

Ulrich Brose, Richard J Williams, and Neo D Martinez. Allometric scaling enhances stability in complex food webs. *Ecology Letters*, 9 (11):1228–1236, 2006.

Sergey V Buldyrev, Roni Parshani, Gerald Paul, H Eugene Stanley, and Shlomo Havlin. Catastrophic cascade of failures in interdependent networks. *Nature*, 464(7291):1025, 2010.

Alyssa R Cirtwill, Tomas Roslin, Claus Rasmussen, Jens Mogens Olesen, and Daniel B Stouffer. Between-year changes in community composition shape species' roles in an arctic plant–pollinator network. *Oikos*, 2018.

José M Costa, Jaime A Ramos, Sérgio Timóteo, Luis P da Silva, Ricardo S Ceia, and Ruben H Heleno. Species activity promote the stability of fruit-frugivore interactions across a five-year multilayer network. *bioRxiv*, page 421941, 2018.

Peter A Cotton. Coevolution in an amazonian hummingbird-plant community. *Ibis*, 140(4):639–646, 1998.

Michael M Danziger, Louis M Shekhtman, Amir Bashan, Yehiel Berezin, and Shlomo Havlin. Vulnerability of interdependent networks and networks of networks. In *Interconnected Networks*, pages 79–99. Springer, 2016.

Jennifer A Dunne and Richard J Williams. Cascading extinctions and community collapse in model food webs. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1524):1711–1723, 2009.

Frank N Egerton. *Roots of ecology: antiquity to Haeckel*. Univ of California Press, 2012.

Anna Eklöf, Ute Jacob, Jason Kopp, Jordi Bosch, Rocío Castro-Urgal, Natacha P Chacoff, Bo Dalsgaard, Claudio Sassi, Mauro Galetti, Paulo R Guimaraes, et al. The dimensionality of ecological networks. *Ecology Letters*, 16(5):577–583, 2013.

Colin Fontaine, Paulo R Guimarães, Sonia Kéfi, Nicolas Loeuille, Jane Memmott, Wim H van Der Putten, Frank JF van Veen, and Elisa Thébault. The ecological and evolutionary implications of merging different types of networks. *Ecology Letters*, 14(11):1170–1181, 2011.

Miguel A Fortuna, Daniel B Stouffer, Jens M Olesen, Pedro Jordano, David Mouillot, Boris R Krasnov, Robert Poulin, and Jordi Bascompte. Nestedness versus modularity in ecological networks: two sides of the same coin? *Journal of Animal Ecology*, 79(4):811–817, 2010.

M. P. Gaiarsa and P Guimaraes Jr. Interaction strength promote robustness in mutualistic networks against cascading effects. *Scientific Reports*, in press.

Mauro Galetti, Roger Guevara, Marina C Côrtes, Rodrigo Fadini, Sandro Von Matter, Abraão B Leite, Fábio Labecca, Thiago Ribeiro, Carolina S Carvalho, Rosane G Collevatti, et al. Functional extinction of birds drives rapid evolutionary changes in seed size. *Science*, 340 (6136):1086–1090, 2013.

David García-Callejas, Roberto Molowny-Horas, and Miguel B Araújo. Multiple interactions networks: towards more realistic descriptions of the web of life. *Oikos*, 127(1):5–22, 2018.

Ewen Georgelin and Nicolas Loeuille. Evolutionary response of plant interaction traits to nutrient enrichment modifies the assembly and structure of antagonistic-mutualistic communities. *Journal of Ecology*, 104(1):193–205, 2016.

Oscar Godoy, Ignasi Bartomeus, Rudolf P Rohr, and Serguei Saavedra. Towards the integration of niche and network theories. *Trends in Ecology and Evolution*, 2018.

Carlos Gracia-Lázaro, Laura Hernández, Javier Borge-Holthoefer, and Yamir Moreno. The joint influence of competition and mutualism on the biodiversity of mutualistic ecosystems. *Scientific Reports*, 8 (1):9253, 2018.

Dominique Gravel, François Massol, and Mathew A Leibold. Stability and complexity in model meta-ecosystems. *Nature Communications*, 7:12457, 2016.

Jessica L Green, Alan Hastings, Peter Arzberger, Francisco J Ayala, Kathryn L Cottingham, Kim Cuddington, Frank Davis, Jennifer A Dunne, Marie-Josee Fortin, Leah Gerber, et al. Complexity in ecology and conservation: mathematical, statistical, and computational challenges. *AIBS Bulletin*, 55(6):501–510, 2005.

Paulo R Guimarães Jr, Mathias M Pires, Pedro Jordano, Jordi Bascompte, and John N Thompson. Indirect effects drive coevolution in mutualistic networks. *Nature*, 550(7677):511, 2017.

Yann Hautier, Pascal A Niklaus, and Andy Hector. Competition for light causes plant biodiversity loss after eutrophication. *Science*, 324(5927):636–638, 2009.

Gareth P Hempson, Sally Archibald, and William J Bond. A continent-wide assessment of the form and intensity of large mammal herbivory in africa. *Science*, 350(6264):1056–1061, 2015.

Inga Hiiesalu, MAARJA OePIK, Madis Metsis, Liisa Lilje, John Davison, Martti Vasar, Mari Moora, Martin Zobel, Scott D Wilson, and Meelis Paertel. Plant species richness belowground: higher richness and new patterns revealed by next-generation sequencing. *Molecular Ecology*, 21(8):2004–2016, 2012.

Crawford S Holling. Some characteristics of simple types of predation and parasitism. *The Canadian Entomologist*, 91(7):385–398, 1959.

Xuqing Huang, Jianxi Gao, Sergey V Buldyrev, Shlomo Havlin, and H Eugene Stanley. Robustness of interdependent networks under targeted attack. *Physical Review E*, 83(6):065101, 2011.

Matthew C Hutchinson, Edgar Fernando Cagua, and Daniel B Stouffer. Cophylogenetic signal is detectable in pollination interactions across ecological scales. *Ecology*, 98(10):2640–2652, 2017.

Pedro Jordano. Sampling networks of ecological interactions. *Functional Ecology*, 30(12):1883–1893, 2016.

Tyler R Kartzinel, Patricia A Chen, Tyler C Coverdale, David L Erickson, W John Kress, Maria L Kuzmina, Daniel I Rubenstein, Wei Wang, and Robert M Pringle. Dna metabarcoding illuminates dietary niche partitioning by african large herbivores. *Proceedings of the National Academy of Sciences*, 112(26):8019–8024, 2015.

Sonia Kéfi, Vincent Miele, Evie A Wieters, Sergio A Navarrete, and Eric L Berlow. How structured is the entangled bank? the surprisingly simple organization of multiplex ecological networks leads to

increased persistence and resilience. *PLoS Biology*, 14(8):e1002527, 2016.

Anne Kempel, Martin Schädler, Thomas Chrobock, Markus Fischer, and Mark van Kleunen. Tradeoffs associated with constitutive and induced plant resistance against herbivory. *Proceedings of the National Academy of Sciences*, 108(14):5685–5689, 2011.

Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P Gleeson, Yamir Moreno, and Mason A Porter. Multilayer networks. *Journal of Complex Networks*, 2(3):203–271, 2014.

Idaline Laigle, Isabelle Aubin, Christoph Digel, Ulrich Brose, Isabelle Boulangeat, and Dominique Gravel. Species traits as drivers of food web structure. *Oikos*, 127(2):316–326, 2018.

Jose B Lanuza, Ignasi Bartomeus, and Oscar Godoy. Opposing effects of floral visitors and soil conditions on the determinants of competitive outcomes maintain species diversity in heterogeneous landscapes. *Ecology Letters*, 21(6):865–874, 2018.

Terence M Laverty and RC Plowright. Competition between hummingbirds and bumble bees for nectar in flowers of impatiens biflora. *Oecologia*, 66(1):25–32, 1985.

John H Lawton. Are there general laws in ecology? *Oikos*, pages 177–192, 1999.

Stan L Lindstedt, Brian J Miller, and Steven W Buskirk. Home range, time, and body size in mammals. *Ecology*, 67(2):413–418, 1986.

Run-Ran Liu, Daniel A Eisenberg, Thomas P Seager, and Ying-Cheng Lai. The "weak" interdependence of infrastructure systems produces mixed percolation transitions in multilayer networks. *Scientific Reports*, 8(1):2111, 2018.

Gianalberto Losapio, Francisco I Pugnaire, Michael J O'Brien, and Christian Schöb. Plant life history stage and nurse age change the development of ecological networks in an arid ecosystem. *Oikos*, 2018.

Robert MacArthur and Richard Levins. The limiting similarity, convergence, and divergence of coexisting species. *The American Naturalist*, 101(921):377–385, 1967.

John L Maron and Montserrat Vilà. When do herbivores affect plant invasion? evidence for the natural enemies and biotic resistance

hypotheses. *Oikos*, 95(3):361–373, 2001.

Robert M May. Will a large complex system be stable? *Nature*, 238 (5364):413, 1972.

Robbie A McDonald. Resource partitioning among british and irish mustelids. *Journal of Animal Ecology*, 71(2):185–200, 2002.

Brian J McGill, Brian J Enquist, Evan Weiher, and Mark Westoby. Rebuilding community ecology from functional traits. *Trends in Ecology and Evolution*, 21(4):178–185, 2006.

Millennium Ecosystem Assessment. *Ecosystems and human well-being.* Island Press, Washington, D.C., USA, 2005.

Michaël AJ Moens, Gediminas Valkiūnas, Anahi Paca, Elisa Bonaccorso, Nikolay Aguirre, and Javier Pérez-Tris. Parasite specialization in a unique habitat: hummingbirds as reservoirs of generalist blood parasites of andean birds. *Journal of Animal Ecology*, 85(5): 1234–1245, 2016.

Mohammad Sadegh Norouzzadeh, Anh Nguyen, Margaret Kosmala, Alexandra Swanson, Meredith S Palmer, Craig Packer, and Jeff Clune. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences*, page 201719367, 2018.

Mark Novak, Justin D Yeakel, Andrew E Noble, Daniel F Doak, Mark Emmerson, James A Estes, Ute Jacob, M Timothy Tinker, and J Timothy Wootton. Characterizing species interactions to understand press perturbations: what is the community matrix? *Annual Review of Ecology, Evolution, and Systematics*, 47:409–432, 2016.

Jens M Olesen, Jordi Bascompte, Yoko L Dupont, and Pedro Jordano. The modularity of pollination networks. *Proceedings of the National Academy of Sciences*, 104(50):19891–19896, 2007.

Jens M Olesen, Jordi Bascompte, Yoko L Dupont, Heidi Elberling, Claus Rasmussen, and Pedro Jordano. Missing and forbidden links in mutualistic networks. *Proceedings of the Royal Society B: Biological Sciences*, page rspb20101371, 2010.

Robert T Paine. Road maps of interactions or grist for theoretical development? *Ecology*, 69(6):1648–1654, 1988.

Mercedes Pascual, Jennifer A Dunne, et al. *Ecological networks: linking structure to dynamics in food webs*. Oxford University Press, 2006.

Shai Pilosof, Mason A Porter, Mercedes Pascual, and Sonia Kéfi. The multilayer nature of ecological networks. *Nature Ecology and Evolution*, 1:0101, 2017.

Shai Pilosof, Qixin He, Kathryn E Tiedje, Shazia Ruybal-Pesantez, Karen Day, and Mercedes Pascual. Competition for hosts modulates vast antigenic diversity to generate persistent strain structure in plasmodium falciparum. *bioRxiv*, page 406546, 2018.

Jorrit H Poelen, James D Simons, and Chris J Mungall. Global biotic interactions: an open infrastructure to share and analyze species-interaction datasets. *Ecological Informatics*, 24:148–159, 2014.

Timothée Poisot, Elsa Canard, David Mouillot, Nicolas Mouquet, and Dominique Gravel. The dissimilarity of species interaction networks. *Ecology Letters*, 15(12):1353–1361, 2012.

Timothée Poisot, Benjamin Baiser, Jennifer A Dunne, Sonia Kéfi, Francois Massol, Nicolas Mouquet, Tamara N Romanuk, Daniel B Stouffer, Spencer A Wood, and Dominique Gravel. mangal–making ecological network analysis simple. *Ecography*, 39(4):384–390, 2016.

Lauren C Ponisio, Marilia P Gaiarsa, and Claire Kremen. Opportunistic attachment assembles plant–pollinator networks. *Ecology Letters*, 20(10):1261–1272, 2017.

Navin Ramankutty, Amato T Evan, Chad Monfreda, and Jonathan A Foley. Farming the planet: 1. geographic distribution of global agricultural lands in the year 2000. *Global Biogeochemical Cycles*, 22(1), 2008.

Saulo DS Reis, Yanqing Hu, Andrés Babino, José S Andrade Jr, Santiago Canals, Mariano Sigman, and Hernán A Makse. Avoiding catastrophic failure in correlated networks of networks. *Nature Physics*, 10(10):762, 2014.

Jay A Rosenheim. Source–sink dynamics for a generalist insect predator in habitats with strong higher-order predation. *Ecological Monographs*, 71(1):93–116, 2001.

VHW Rudolf and Kevin D Lafferty. Stage structure alters how complexity affects stability of ecological networks. *Ecology Letters*, 14 (1):75–79, 2011.

Peter F Sale. *The Ecology of Fishes on Coral Reefs*. Elsevier, 2013.

Pälvi Salo, Erkki Korpimäki, Peter B Banks, Mikael Nordström, and Chris R Dickman. Alien predators are more dangerous than native predators to prey populations. *Proceedings of the Royal Society B: Biological Sciences*, 274(1615):1237–1243, 2007.

Alix Sauve, Elisa Thébault, Michael JO Pocock, and Colin Fontaine. How plants connect pollination and herbivory networks and their contribution to community stability. *Ecology*, 97(4):908–917, 2016.

JS Singh and PS Yadava. Seasonal variation in composition, plant biomass, and net primary productivity of a tropical grassland at kurukshetra, india. *Ecological Monographs*, 44(3):351–376, 1974.

F Gary Stiles. Possible specialization for hummingbird-hunting in the tiny hawk. *The Auk*, pages 550–553, 1978.

Elisa Thébault and Colin Fontaine. Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science*, 329(5993):853–856, 2010.

John N Thompson. *The geographic mosaic of coevolution*. University of Chicago Press, Chicago, IL, 2005.

Sérgio Timóteo, Marta Correia, Susana Rodríguez-Echeverría, Helena Freitas, and Ruben Heleno. Multilayer networks reveal the spatial structure of seed-dispersal interactions across the great rift landscapes. *Nature Communications*, 9(1):140, 2018.

Hirokazu Toju and Yuki G Baba. Dna metabarcoding of spiders, insects, and springtails for exploring potential linkage between above-and below-ground food webs. *Zoological Letters*, 4(1):4, 2018.

Diego P Vázquez, William F Morris, and Pedro Jordano. Interaction frequency as a surrogate for the total effect of animal mutualists on plants. *Ecology Letters*, 8(10):1088–1094, 2005.

B

SUPPLEMENTARY INFORMATION S1

Supporting information for Chapter 1: Unmasking structural patterns in incidence matrices: an application to ecological data.

B.1 Supplementary Methods

B.1.1 Null model overlap

An important aspect to consider when comparing the performance of multiple null models is how different the structure of the data generated by them is from the original data (Joppa et al., 2010). The reason is that there is generally a number of shared links between the empirical adjacency matrix and the generated data, and these shared links may be crucial to explain the observed non-random patterns (Joppa and Williams, 2011). In other words, the number of shared links approximates the amount of variation between the data generated by the null model and the original adjacency matrix, and therefore is indicative of the "effective degrees of freedom" of a particular randomization scheme. We define the degree of overlap between the structure of an adjacency matrix and a particular ensemble of randomized networks as the average number of empirical links conserved after the randomization process. That is, given the adjacency matrix A and a particular randomized matrix A^*, we estimate the number of shared links between them as

$$\Theta(A^*) = \sum_{ij} \left(A_{ij} - \frac{|A_{ij} - A^*_{ij}|}{2} \right). \tag{10}$$

The typical overlap between an adjacency matrix and a random ensemble can be defined then as the mean proportion of links shared between the former and each of the randomized matrices.

B.1.2 Estimating food-web phylogenies

To quantify any evolutionary signal underlying food-web structure, we first needed to generate phylogenies for the different species. To do so, we started by taxonomically classifying all species according to NCBI database (http://www.ncbi.nlm.nih.gov/) by means of the *classification* function in the R package *taxize* (Chamberlain and Szöcs,

2013). Using this information, we could obtain the cladograms corresponding to the species' taxonomy using the *as.phylo.formula* function from *ape* (Paradis et al., 2004). While doing so, we considered all species with an indefinite taxonomic classification as outgroups (e.g., moss cells and unidentified detritus). We then calibrated the resultant trees based on published data of the actual divergence time between species. Whenever possible, we dated the ancestral nodes of the cladograms—the most recent common ancestor of two given taxa—according to Hedges et al. (2006, 2015), which is a database of published molecular divergence times for a large number of species (<50000). Finally, the age of all remaining undated ancestral nodes was estimated according to the branch length adjustment algorithm *bladj* (Webb et al., 2008), which evenly sets the undated nodes between dated ones.

B.1.3 *Generating structured food-web structures*

To generate structured food-web data, we chose to use the niche model presented by Williams and Martinez 2000. Given a set of n species, the niche model simulates the structure of a food web by assigning a random 'niche value' from the interval $[0, 1]$ to all species that determines who is eaten by whom. In particular, every species i with niche value k_i is set to consume any species that have a niche value falling in a particular range r_i randomly centered between $r_i/2$ and k_i, where r_i is drawn from a beta distribution with $\alpha = 1$ and expected value $2c$, and c is the desired connectance for the food web.

B.1.4 *Generating structured species assemblages*

To generate nested species assemblages, we follow a twofold process. First, given a number of rows n, columns m and connectance c, we add interactions $m_ij = 1$ to generate a perfectly nested matrix M. In particular, we fill the matrix following Patterson and Atmar 1986 to obtain a perfectly nested structure. Second, we add noise to the interaction matrix. To do so, we go through every element m_{ij} of the matrix and switch it with a given probability p (i.e. we change every element $m_{ij} = 1$ to $m_{ij} = 0$ with probability p, and vice-versa)

B.2 *Supplementary Results*

B.2.1 *Food webs and overlap*

We evaluated the degree of overlap between the empirical food-webs and the data generated by the null models to ensure that the observed differences in the motif representation were not a consequence of the strengths of the imposed constrains (Rohr et al., 2014). That is,

we wanted to verify that the aforementioned differences were not due to the number of shared links between the empirical and random structures but instead arose from the intrinsic properties of the adopted null hypotheses. We observed that the overlap between the empirical networks and the random ensemble representing the uninformed null model ranged from 39% to 65%. On the other side, the degree of overlap was consistently lower for the data generated by phylogenetically-informed null model, ranging from 53% to 73% when we used the estimated probabilities according to predator's diet and from 48% to 69% when we adopted the alternative prey's consumers perspective.

To verify that the these overlap differences were not the responsible for the different motif representations, we evaluated the trajectories followed by the z-scores when progressively randomizing a given food web. That is, we quantified the motif over- and underrepresentation according to random ensembles presenting different degrees of overlap. Using these trajectories, we studied the motif representation as a function of the average number of links shared by the empirical food webs and the data generated by the different null models. We observed that the motif composition of a particular food web was very robust to changes accounting for the species' phylogenetic relationships. On the other side, this empirical motif composition was very sensitive to the uninformed randomization, showing a very different pattern relative to the phylogenetically-informed one. Figure 18 shows an example of the pattern found for the motif describing exploitative competition in a particular network.

Finally, we also confirmed this result by artificially increasing the overlap shown by the data generated by the uninformed null model relative to the phylogenetically-informed versions. In order to ensure that the differences observed between the motif representations obtained using those null models were not due to their different degrees of overlap, we artificially increased the overlap of the uninformed null model to match the overlap of the informed ones. To do so, we used a Markov chain Monte Carlo switching algorithm. Specifically, given an adjacency matrix A representing a food-web structure and a null hypothesis described by the ensemble of randomized networks $\langle A^* \rangle$, we progressively reallocated the links of each A^* according to A until obtaining a particular overlap between them. In this process, for example, two links $i \leftarrow j$ and $l \leftarrow m$ can become $i \leftarrow m$ and $l \leftarrow j$, provided that the overlap between A^* and A either increases or stays the same. Using this algorithm, we generated random ensembles describing the uninformed and the phylogenetically-informed null hypothesis with the same degree of overlap relative to the empirical food webs. Figure 19 shows an example of the results obtained when arti-

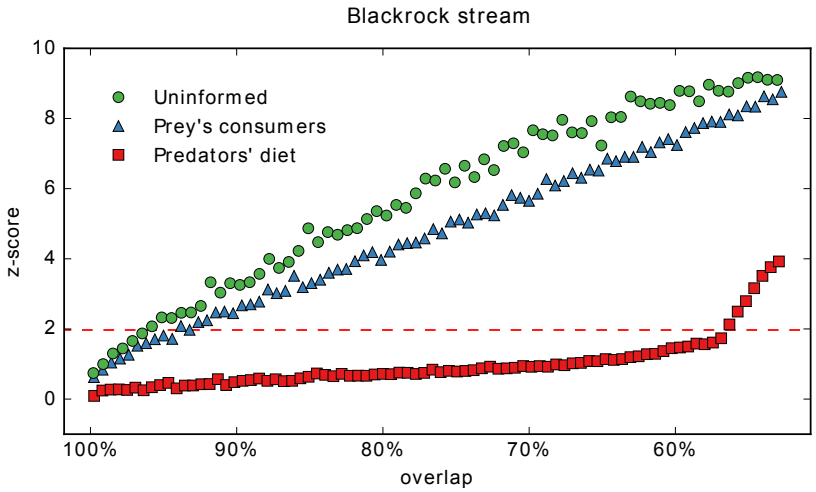


Figure 18: The relationship between the motif representation of a simple food chain and the degree of overlap of the data generated by the null models for one of the empirical food webs studied here. The red and blue circles show the trajectories when the randomization accounts for the phylogenetic relationships in predators' diets and prey's consumers, respectively. Similarly, the green circles show the same according to the uninformed null model. The red dotted line indicates the threshold for significance $z = 1.96$.

ficially increasing the overlap of the uninformed null model to match the overlap of the phylogenetically-informed null model based on the predator's diet.

B.2.2 Benchmark testing

The core idea behind benchmark testing null models is to compare their performance when looking at both structured and random data. Specifically, a null model is expected to show a structural pattern to be significantly represented in structured data (low type II error), and unsignificantly represented in random data (low type I error). The uninformed null model used in the main text follow a 'swap' algorithm (Connor and Simberloff, 1979), which has been largely used and tested effective in the past (Milo et al., 2003; Itzkovitz et al., 2004). Therefore, to test the performance of the correlation-informed null model, we studied how an informed null model and a misinformed null model (i.e. informed by a randomized correlation matrix) compared to the uninformed null model in structured and random data. We expected the misinformed null model to show similar results to the uninformed one. Ideally, we instead expected the informed null model to show similar results to the uninformed null model when analyzing random data but very different results when studying structured food webs. In particular, given the right correlation matrix, the informed null model should be able to reproduce the patterns found in structured data.

For the application of the correlation-informed null model on food webs, we first generated 1000 food webs using the algorithm defined in the Supplementary Methods section, and 1000 random matrices. We generated all these matrices randomly picking the number of species $n \in [0, 1]$ and connectance $c \in [0.05, 0.15]$. For each of the structured food webs, we used the niche values for all species to generate a correlation matrix. To do so, we assumed an exponential correlation structure and calculate the correlation matrix using the R package *nlme*. Then, we used an uninformed, a misinformed, a niche-informed null model to analyze the shape of both structured and random food webs (Fig. 20). Following the example in the main text, we studied the representation of all three-species food-web motifs, focusing on whether the null models showed the data to present significant or unsignificant patterns. We found the correlation-informed null model

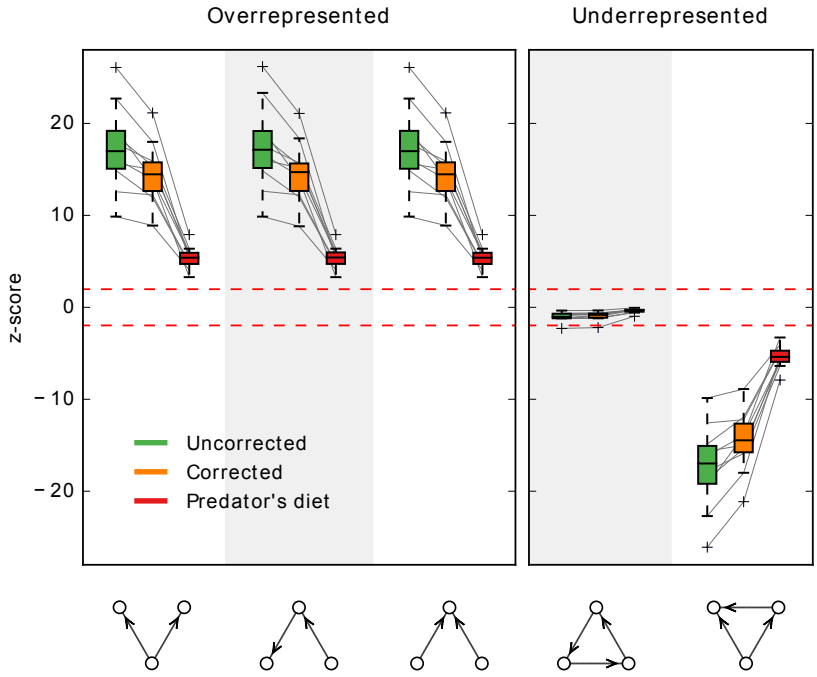


Figure 19: The effect of the phylogenetic relationships between species on the motif representation of a set of food webs. For all motifs, the arrows indicates the transfer of energy from prey to predators. The red dotted line indicates the thresholds for significance $z \leqslant -1.96$ and $z \geqslant 1.96$. The boxes group all food webs, extending from the lower to upper quartile values of the data, with a line at the median (the grey lines connecting the boxes link the motif representation for the same food webs). The green boxes show the motif representation according to the uninformed null model when this has been not constrained. Similarly, the orange boxes show the same when the degree of overlap has been constrained. Finally, the red boxes contain the z-scores for each motif when the null model accounts for the phylogenetic relationships in predators' diets.

to perform almost exactly as the uninformed null model did (Fig. 20). We found the niche-informed null model, on the other side, to explain much better the motif pattern found in structured food webs while not producing any misleading results in randomized structures (low type I error; Fig. 20).

For the second application of the correlation-informed null model, we first generated 1000 species assemblages using the algorithm defined in the Supplementary Methods section, and 1000 random matrices. We generated all these matrices randomly picking the number of rows $n \in [20, 100]$, the number of columns $m \in [20, 100]$, the connectance $c \in [0.05, 0.15]$, and the added noise $p \in [0.1, 0.3]$ (see Supplementary Methods). Following the same procedure as in our previous example, we generate the correlation matrices assuming an exponential correlation structure, using the column order as our similarity measure in this case. Again, we used an uninformed, a misinformed, an informed null model to analyze the shape of both structured and random species assemblages (Fig. 20). Following the example in the main text, we studied the nestedness pattern observed in the simulated species assemblages. As in the test for the food webs, we found

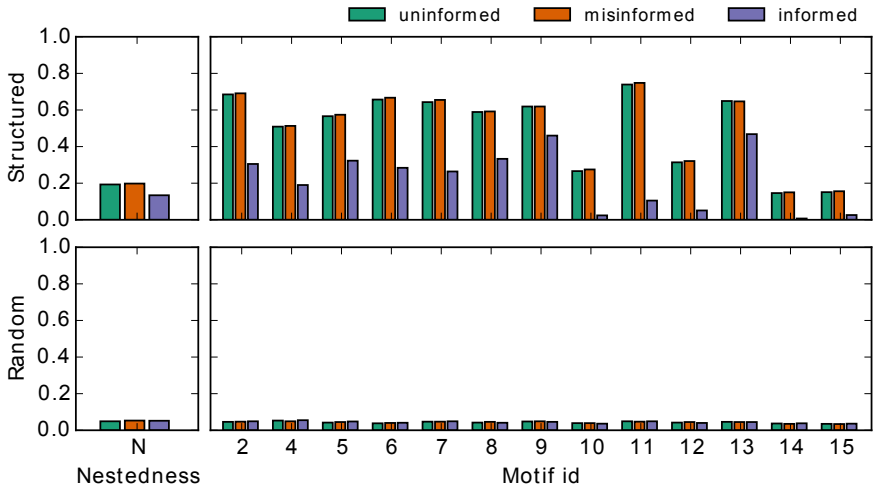


Figure 20: Proportion of times the different null models showed a significant structural pattern in structured and random simulated data. For all cases, we studied the simulated data using three null models: an uninformed, a misinformed and an informed null model. Each bar represents the proportion of times the each model showed the data to present a non-random pattern. The two top panels show the results found for the study of nestedness and the motif representation in structured data, and the bottom panels show the same for random data. The different motif id characterize the distinct isomorphism classes defined by the function '`graph_from_isomorphism_class`' in the R package *igraph*.

the misinformed null model to perform almost exactly as the uninformed null model did (Fig. 20), and the informed null model to better explain the nested pattern found in structured species assemblages while not producing any misleading results in randomized structures (low type I error; Fig. 20).

BIBLIOGRAPHY

Scott A Chamberlain and Eduard Szöcs. taxize: taxonomic search and retrieval in R. *F1000Research*, 2, 2013.

Edward F Connor and Daniel Simberloff. The assembly of species communities: chance or competition? *Ecology*, 60(6):1132–1140, 1979.

S Blair Hedges, Joel Dudley, and Sudhir Kumar. Timetree: a public knowledge-base of divergence times among organisms. *Bioinformatics*, 22(23):2971–2972, 2006.

S Blair Hedges, Julie Marin, Michael Suleski, Madeline Paymer, and Sudhir Kumar. Tree of life reveals clock-like speciation and diversification. *Molecular Biology and Evolution*, 34(4):835–845, 2015.

S Itzkovitz, R Milo, N Kashtan, MEJ Newman, and U Alon. Reply to "comment on 'subgraphs in random networks' ". *Physical Review E*, 70(5):058102, 2004.

Lucas N Joppa and Rich Williams. The influence of single elements on nested community structure. *Methods in Ecology and Evolution*, 2(5):541–549, 2011.

Lucas N Joppa, José M Montoya, Solé Vicente, Jim Sanderson, Stuart L Pimm, et al. On nestedness in ecological networks. *Evolutionary Ecology Research*, 12:35–46, 2010.

Ron Milo, Nadav Kashtan, Shalev Itzkovitz, Mark EJ Newman, and Uri Alon. On the uniform generation of random graphs with prescribed degree sequences. *arXiv preprint cond-mat/0312028*, 2003.

Emmanuel Paradis, Julien Claude, and Korbinian Strimmer. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20(2):289–290, 2004.

Bruce D Patterson and Wirt Atmar. Nested subsets and the structure of insular mammalian faunas and archipelagos. *Biological Journal of the Linnean Society*, 28(1-2):65–82, 1986.

Rudolf P Rohr, Serguei Saavedra, and Jordi Bascompte. On the structural stability of mutualistic systems. *Science*, 345(6195):1253497, 2014.

Campbell O Webb, David D Ackerly, and Steven W Kembel. Phylocom: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics*, 24(18):2098–2100, 2008.

Richard J Williams and Neo D Martinez. Simple rules yield complex food webs. *Nature*, 404(6774):180, 2000.

C

SUPPLEMENTARY INFORMATION S2

Supporting information for Chapter 2: Identifying a common backbone of interactions underlying food webs from different ecosystems.

C.1 Supplementary Results

C.1.1 Structural differences across ecosystems

Using the estimated alignments between food webs, we calculated four dissimilarity matrices $\hat{E}$ containing the pairwise distances between the 411 food webs based on each of the four alignment quality measures defined in the Supplementary Methods section. These dissimilarity matrices allow us to test whether or not food webs from different ecosystems generally share a common structure. To test this, we used a permutational multivariate analysis of variance (PERMANOVA; Anderson, 2001). First, we tested for general differences in the alignment centroids among the different ecosystem types when using the whole dissimilarity matrix (Supplementary Table 2). To avoid aligning networks with very different sizes, we then repeated the test using only a subset of the networks. In particular, we repeated the test considering the biggest possible subset of networks from our dataset, such that the biggest network was at most twice as big as the smallest network (Supplementary Table 2). Finally, we performed a Principal Coordinates Analysis (PCoA) of the whole dissimilarity matrix to visualize the observed differences (Fig. 5 of the main text).

C.1.2 Pairwise comparisons between ecosystems

In this section, we first individually studied the differences between every pair of ecosystem types. Specifically, we compared the 34 estuaries, 87 lakes, 148 marine ecosystems, 88 streams and 54 terrestrial ecosystems using the alignment measure A (Supplementary Table 3). Then, we performed a principal coordinate analysis of every comparison (Supplementary Fig. 21).

C.1.3 Connectance and path likelihood

One of the necessary conditions for the existence of a backbone of interactions is that the best aligned species need to form a connected component. The two measures that we used to test this idea are the

connectance and path likelihood. For every network, we found that the substructures formed by the sets of best aligned species tend to present both a high connectance and path probability when compared to equal sized random subsets of species (Supplementary Fig. 23).

C.1.4 *Links removals: alternative measure for identifying the backbones of interactions*

To compare the backbones of interactions found for every network, we used a different measure of network similarity as an alternative to backbone alignment. Because the backbones are defined by determining the most-overlapped interactions across alignments, it was important that the measure of network similarity not consider differences in the number of species comprising the backbone but only differences in the interactions forming it. In particular, we defined a measure based on "link removals", where the distance between two backbones C and D is calculated as the minimum set of interactions that need to be removed from C and D before they turn into isomorphic structures. To estimate such a distance for every pair of backbones, given two structures made of n links, we recursively try all the potential k-link-removal permutations on both networks until we find isomorphic structures. That is, for one link removal, for example, we would test whether or not there is any combination of one-link removals that produces two isomorphic structures made of $n-1$ links.

For this alternative analysis, we concentrated on the backbones of interactions made up of the six most-overlapped links, since this method is computationally very expensive for large backbones and this is the number of links of the smallest network in our dataset. Comparing the backbones using the measure of link removals, we looked for the two structures that could explain most of the observed backbones (Fig. 27). In particular, we found that these two structures together can explain approximately 60% of the networks' backbones within one 'link removal'. Importantly, the structures found are consistent with the previous results, since they are substructures of the backbones shown in Fig. 7 of the main text.

C.1.5 *Many-link backbones of interactions*

Given the clustering analysis performed for a particular backbone size k, we can assess how distinct the observed clusters are by analyzing the 95% confidence ellipses (see Fig. 7 of the main text). In particular, we looked at the amount of overlap between these ellipses and found that it drastically increased when the backbone size was $k > 15$ (Fig. 28). This could highlight the fact that the noise associated with the shape of the backbones increases with their size, which

is not surprising because the periphery around the backbones seem to differ across networks (Fig. 6 of the main text).

C.1.6 *Backbones of interactions for each ecosystem*

Comparing the backbones of interactions that one would independently find for every ecosystem is nuanced for two main reasons. First, although the existence of backbones of interactions underlying all food webs can be assessed by looking at the alignment transitivity and connectance, we estimated the shapes of such backbones by overlapping the different networks in question. This overlap provided us with an intuitive shape of what the backbones might look like. However, this shape can entail a lot of variability, and one ideally wants to include as many networks as possible to obtain results that are robust to this variability. Focusing on individual ecosystems necessarily limits the number of networks we can use in the analyses, and, therefore, reduces the strength of our results. Second, we do not have the same number of networks for every ecosystem, which also makes the comparison of the resultant backbones in some way uncertain.

Despite these subtle but important distinctions, we repeated the analysis described in the main text for each ecosystem individually. That is, for every ecosystem, we identified every network's backbone of interactions made up of the k most-overlapped links. In this case, we focused only on the backbones made of 6 links, which is the size of the smallest network in our database. For every ecosystem type t, we then aligned the corresponding backbones and generated the dissimilarity matrices E_6^t, where every element $e_{ij|6}$ in these matrices is the optimal alignment cost between the 6-link backbones from any network i and j (Eq. 6 of the main text). Similar to what we found across all ecosystem types, the analysis of these dissimilarity matrices using clustering techniques revealed the existence of two backbones of interactions for any dissimilarity matrix E_6^t (Fig. 29). This implies that two distinct backbones appear regardless of the ecosystem used for the analysis, which could indicate that the backbones found for the entire dataset are general across ecosystems.

To test this idea, we also studied whether or not the clustering found for the entire dataset agrees with the clustering found for the different ecosystems. Using a PERMANOVA analysis, we compared the dissimilarity matrices E_6^t to the results for the clustering analysis of the entire dataset. With the exception of estuary food webs ($p = 0.951$), we found this clustering to be a significant predictor of every dissimilarity matrix E_6^t ($p < 0.01$). This implies that the clustering found for the entire database agrees with the clustering found for the individual ecosystems. The inconclusive results found for the estuarine food webs may relate to an ecological phenomenon or simply be the

result of this ecosystem exhibiting the smallest sample size with only 34 networks.

C.1.7 *Trophic level of the backbones*

For every network, we classified the trophic level h of the species based on whether or not they were basal ($h = 0$), intermediate ($h = 1$) or top species ($h = 2$). Then, for each six-link backbone of interactions, we averaged the trophic level of the species across all networks. We found no significant differences in the trophic levels of the species forming each backbone, with one backbone showing a $h = 0.97 \pm 0.20$ and the other backbone showing a $h = 1.01 \pm 0.22$.

C.1.8 *Compartmentalized structure of food webs*

To measure the degree of compartmentalization across the food webs, we used the quality measure of modularity (Stouffer and Bascompte, 2011). To calculate this, we used the algorithm presented by Leicht, E. A. and Newman M. E. J. (Leicht and Newman, 2008), which estimates both the measure of modularity and the optimal number of modules for any given network. We found that almost every network of our dataset showed some degree of compartmentalization, since the networks in our dataset have, on average, 3.45 ± 1.18 modules. In particular, these networks show an average modularity of 0.25 ± 0.10, with only two networks showing null modularity.

C.2 *Supplementary Methods*

C.2.1 *Alignment algorithm*

Given two food webs, we presented an alignment algorithm that aims to find the best matching of species in one network to species in the other. Starting from a random alignment between these food webs, we use a stochastic optimization algorithm to progressively modify this initial alignment and minimize an alignment cost function that increases when species playing similar ecological roles are paired. During the course of our study, we identified three crucial aspects regarding the alignment cost function that one needs to consider: (1) a pairwise measure of role similarity between species, (2) the degree of the alignment, and (3) the contribution of unpaired species.

C.2.1.1 *Pairwise similarity between species roles*

As noted in the main text, we use the definition based on the idea of network motifs (Stouffer et al., 2012)—the unique n-species subnetworks describing all patterns of interactions between n species—to measure the structural role of species in our dataset. For example,

considering two-species food-web motifs would imply studying the distinct subnetworks $a \leftarrow b$ and $c \leftrightarrow d$. The first motif $a \leftarrow b$ represents species a consuming species b and defines two unique positions: the consumer and resource. On the other hand, the second motif $c \leftrightarrow d$ describes two different species consuming each other and represents a single unique position, since c and d are indistinguishable. Following this scheme, species of any given food web could be characterized by a "motif-role profile" accounting for the number of times that they appear in each of the three unique positions of the two-species motifs. The resolution of this motif-role profile can be extended to also consider three-species food-web motifs, which define 13 additional patterns of interaction and 30 new unique positions. As a result, the motif-role profile of any species can be defined as $\vec{c}_a = \{c^2_{a1}, c^2_{a2}, c^2_{a3}, c^3_{a1}, \ldots, c^3_{a30}\}$, where all c^2_{aj} and c^3_{ak} define the frequencies that any species a appears in the two-species-motif positions j and the three-species-motif positions k, respectively. The pairwise role similarity between any two species can then be calculated using those motif-role profiles and by means of Eq. 4 in the Methods section of the main text. Other motifs sizes could also be considered if desired; however, there are 199 four-species motifs and—given the alignment strategy described below—such additional information would likely be redundant and computationally inefficient.

C.2.1.2 *Degree of the alignment*

Following the definition of the pairwise role similarity, we can now compare the roles of individual species from any two food webs A and B; given an alignment between those networks, we can also compute the overall similarity between them. In the paper, we described two possible strategies for aligning networks: pairing species to species and pairing species' neighbors to species' neighbors. The former strategy is characterized by optimizing Eq. 5 in the Methods section of the main text. Notably, this cost function would focus on minimizing the sum across every species-species pairings and might ignore the actual structure of food webs (Supplementary Fig. 30). Instead, the latter strategy is characterized by optimizing Eq. 6 in the Methods section of the main text and pairs up species based on the motif-role profile of their neighbors (Supplementary Fig. 30).

The advantage of pairing species' neighbors to species' neighbors over pairing species to species can be understood using the example presented in Supplementary Fig. 30. In this example, we have two simple food webs aligned as $\{(A, a), (B, b), (C, c), (D, d), (\emptyset, e)\}$, where every element (i, j) represents the pairing between any given two species i and j from each of the networks. Although this alignment is optimal according to any of the two strategies presented here, there are other possible alignments that are as good. In particular,

pairing species to species we would find 4 indistinguishable alignments (Supplementary Fig. 31). The reason is that the cost function associated to such strategy will reach a minimum as long as the individual species' roles match. Therefore, the pairing (A, a) will be as probable as the pairing (A, e) and independent from the pairing of the neighbors of A. In contrast, pairing species' neighbors to species' neighbors only implies finding 2 indistinguishable results because the cost of pairing (A, a) will depend on the pairing of the neighbors of A (Supplementary Fig. 31).

C.2.1.3 *Contribution of unpaired species*

The first aspect to consider regarding the contribution of unpaired species is the penalty ξ_x used in Eq. 6. This penalty could be defined as simply the number of neighbors that have not been paired to any species. However, we are interested in allowing species to remain unpaired if we cannot find them a proper match, and this approach would over-penalize non-pairing. Instead, we define this penalty for the alignment $x = (a, b)$ based on the unpaired neighbors of the species with a higher degree (i.e. the non-paired neighbors of species a when $n_a \geqslant n_b$ and the non-paired neighbors of b, otherwise). Following this, we can rewrite the term ξ_x as $\xi_{(a,b)} = \max\left(k_{x_\alpha}, k_{x_\beta}\right) \times (1 - \varepsilon)$, where k_{x_α} (and k_{x_β}) is the number of neighbors of a (and b) that are not paired with a neighbor of a (and b), and ε is the default penalty associated with an individual unpaired species.

Finally, we also need to specify the default contribution to the cost function associated with species that are not paired (characterized by ε in Eq. 5 and 6 of the main text). Since the algorithm used here allows species from either network to remain unpaired, this contribution defines an important threshold at which we permit non-pairing. Although the value of ε is somewhat arbitrary, it needs to draw a "middle point" between equivalent and opposed motif-role profiles (see Eq. 4 in the Methods), penalizing unpaired species in order to incentivize the alignment between networks but avoiding the pairing of uncorrelated species. We set the contribution of unpaired species to $\varepsilon = 0$ because this represents null correlation according to the measure of pairwise role similarity used here.

C.2.2 *Alignment quality measures*

In order for the alignments to be comparable across of networks, we also needed a size-independent measure of how good such alignment is. Finding this measure is not trivial since some motifs might only appear in big networks, artificially decreasing the overall similarity

between what could be similar structures. Here we proposed four different measures of alignment quality:

- **Measure A.** First, we considered the measure described by Eq. 7 in the Methods section of the main text, which is a normalized version of Eq. 5.

- **Measure B.** The second measure that we define is a normalized version of Eq. 6 from the main text. Given the best alignment $\hat{\lambda}$ between webs A and B, the quality of the alignment could be characterized by

$$\hat{e}_{AB}(\hat{\lambda}) = \frac{1}{N} \sum_{x \in \hat{\lambda}} \frac{1}{M_x} \left(\sum_{(\alpha,\beta) \in \hat{\lambda}_x} \left(1 - \rho\left(\alpha, \beta\right) \right) + \min\left(k_{x_\alpha}, k_{x_\beta}\right) \right), \tag{11}$$

  where $\rho(a, \emptyset) = \rho(\emptyset, b) = 1$, M_x is the number of neighbors of species a paired with neighbors of species b, and N is the total number of matches between one species from A and one species from B. Notice that this alignment quality normalizes all species' neighbors pairings while ignoring species that have not been paired.

- **Measure C.** We also consider a different measure of alignment quality that is not based on the ecological role of the species and does not use the pairwise similarity defined above. This is the proportional link overlap between aligned networks. Given an alignment between two food webs, the link overlap percentage tells us what percentage of links in one web are also present in the other one. Notice that if two networks A and B have a different number of links, an alignment between them can present two measures of the link overlap: the overlap of A relative to B and the overlap of B relative to A. For example, given the alignment presented in Supplementary Fig. 30, the red network presents a 100% link overlap relative to the blue network and the blue networks presents a 75% link overlap with the red one. Since we are interested in a size independent measure of alignment quality, we focus here on the link overlap of the small network relative to the big one.

- **Measure D.** Given the alignment between two networks A and B, the last measure is based around the idea of removing species for which the alignment algorithm could not find them a proper match. That is, we reduced the aligned networks so that only paired species were considered, and then we recalculated measure A. Doing this, we decrease the effect of the existence of species with different motif-role profiles only due to the size of

the networks. For example, following the alignment presented in Supplementary Fig. 30, species *e* would be removed and the resultant networks would present a perfect alignment. Notice that this measure differs from the first measure because we now remove the motif formed by species e, d and c from the analysis; therefore, the post-removal motif-role profiles of every species-species pairing now perfectly match.

C.2.3 *Algorithm tests*

C.2.3.1 *Effect of link and size differences*

We performed two tests for the alignment algorithm and alignment quality measures used in this study. In particular, we studied the effect of link and size differences on a random subset of networks. This is important because we need to check that the alignment quality measures are sensitive to changes in the structure of the networks but neutralize the effect of size difference between the aligned networks. First, we progressively randomized each of the selected networks—shifting them from the empirical networks to random graphs—and aligned the resultant randomized networks to their original versions (Supplementary Fig. 32). The alignment of a network with itself should present a perfect match; therefore, the alignment of a network with a randomized version of itself should highlight the differences in the way links are distributed within networks. Second, we performed a similar test to analyze how size differences alter the alignment between networks. To do so, we randomly removed a fraction of the species of each network and aligned the resultant networks to their original versions (Supplementary Fig. 32). In this case, the alignment of a network with one of its substructures should only highlight potential differences due to network size.

There are two crucial aspects of the tests that are important to analyze. First, the alignment quality measures should converge to a similar value when comparing the different empirical networks to random networks. This seems to be true in all the cases except for measure C, which clearly scales the network dissimilarity value with network size. This is not necessarily a problem for comparing differently sized networks; however, it will certainly add noise to the overall results. The second aspect to consider is the effect of size difference between networks when comparing identical structures. Ideally, the alignment quality measures should be able to neutralize this effect. We notice that measures C and D do a much better job neutralizing such effect than measures A and B. Nevertheless, we also notice that the effect of size is specially relevant when removing random species from small networks. The reason is because it is easier to break those networks into isolated links when randomly removing species. When focusing

on the biggest networks of the test, one can notice that the alignment quality measures actually do a much better job neutralizing the effect of size.

C.2.3.2 *Alignment variability*

Another important aspect to assess regarding the alignment algorithm is the number of possible optimal alignments between two given networks. That is, the amount of variability found across alignments (see example in Supplementary Fig. 30). This is important because we want to know whether or not we need to align the networks multiple times in order to validate our results. Here, we aligned every pair of the networks used in the previous section 100 times and estimated the typical number of pairings of all species. For any given network, we found that the average number of pairings of a species is 2.03 ± 1.88, indicating that aligning networks only once is a reasonable approximation (Supplementary Fig. 30). In contrast, we found that this number is 20.52 ± 10.82 when aligning networks at random. As expected, however, we also found that the number of pairings increases when aligning bigger networks (Supplementary Fig. 30). Intuitively, this emerges since larger networks have a greater probability of including trophically identical species as well as species that participate in just a single interaction.

C.3 Supplementary Figures and Tables

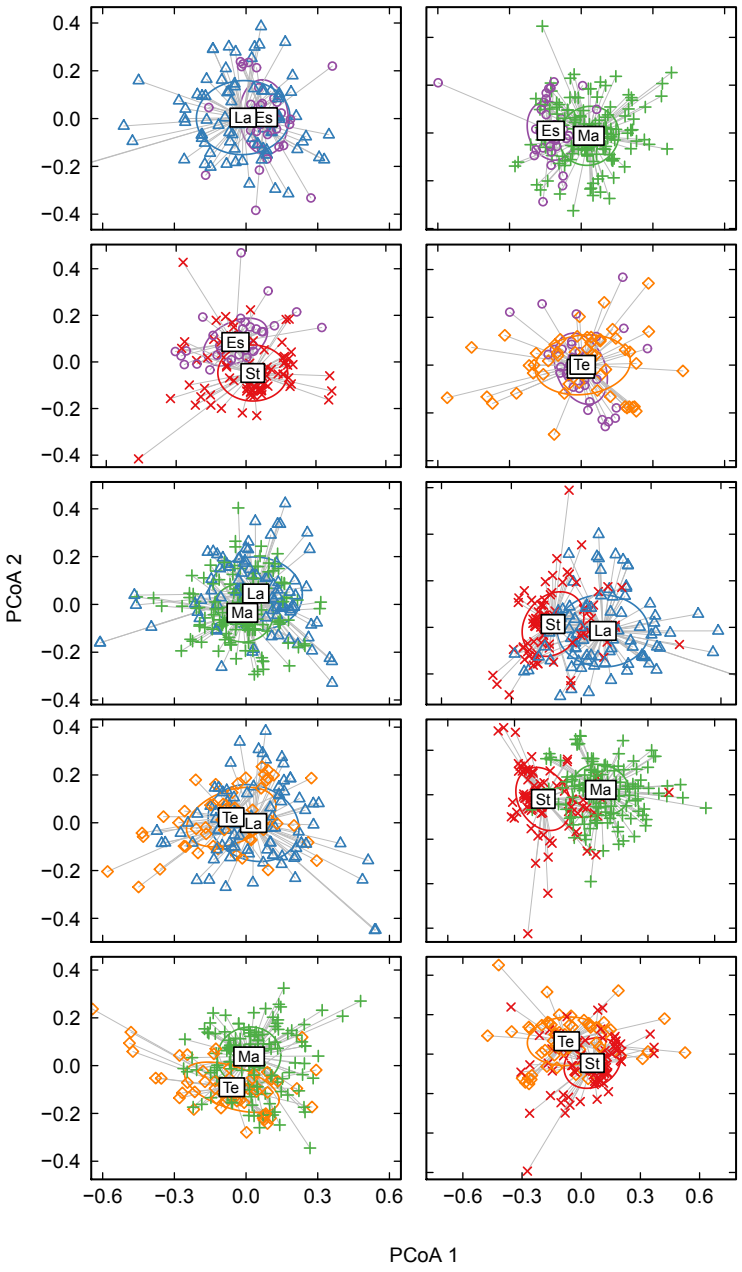


Figure 21: Principal coordinate analysis for every pair of ecosystems. We used the alignment quality measure A, which is the measure used in the paper. As in Supplementary Fig. 2 of the main text, each different color and symbol characterizes the group of networks from estuaries (Es), lakes (La), marine (Ma), streams (St) and terrestrial (Te) ecosystems. The ellipses represent the 1 standard deviation ellipses about the group medians.

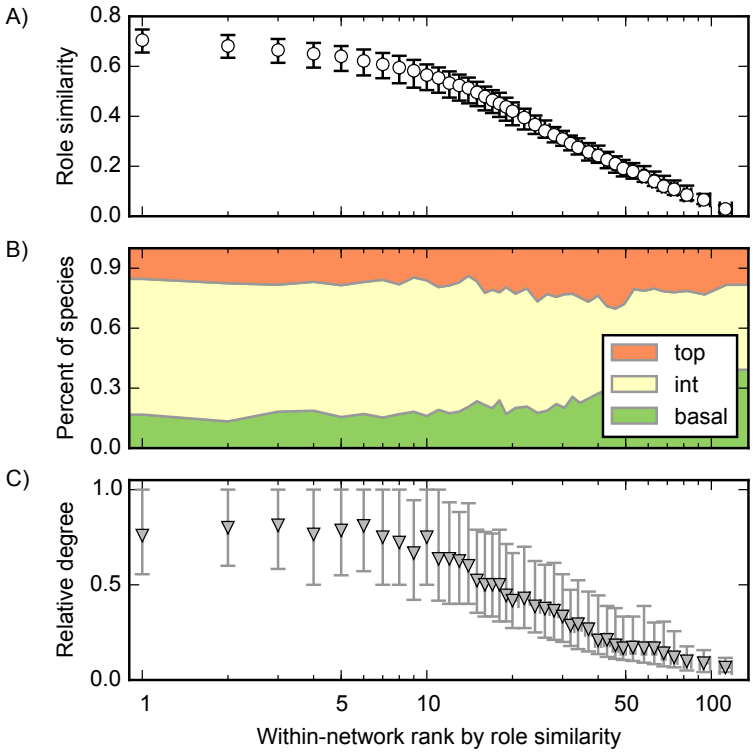


Figure 22: Ranking of species from our dataset of 411 food webs based on the average similarity between their role and the roles of the species to which they are paired across all 84255 alignments. The top panel (A) shows the observed role similarity for all the species. The middle panel (B) describes the proportion of times that the ranked species are at top (i.e. not feeding on any species in the network), basal (i.e. not being consumed by any species in the network), and intermediate (i.e. feeding on and being consumed by other species in the web) trophic levels. The bottom panel (C) shows the relative degree of the ranked species. In all panels, every point indicates the median across at least 250 species with the exception of the last point which is the median across 30 species, and the bars characterize the inter quartile range.

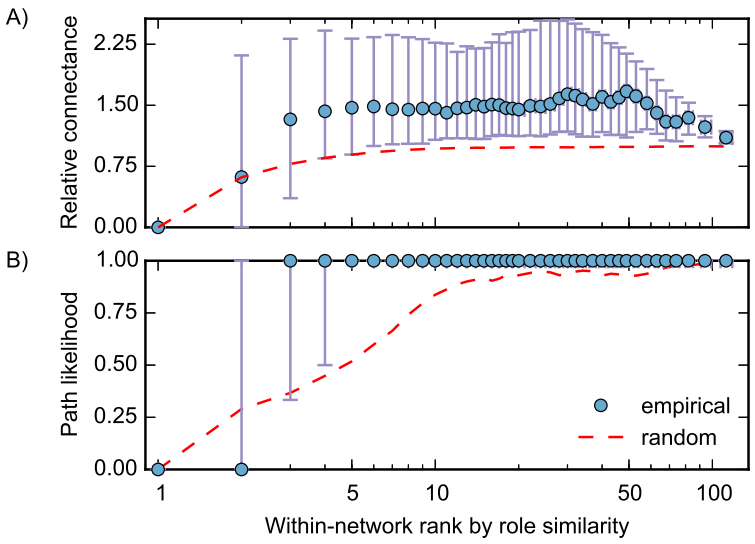


Figure 23: Analysis of the connectance and path probability of the best aligned species of each network. The ranking of species is the same as the one presented in Supplementary Fig. 22. For every value of x, the blue circles represent (A) the relative connectance (connectance relative to the original network) and (B) path probability of the x best aligned species of the ranking. The red dotted lines characterize the relative connectance and path probability expected for a random subset of x species. In both panels, every point indicates the median across at least 250 species with the exception of the last point which is the median across 30 species, and the bars characterize the inter quartile range.

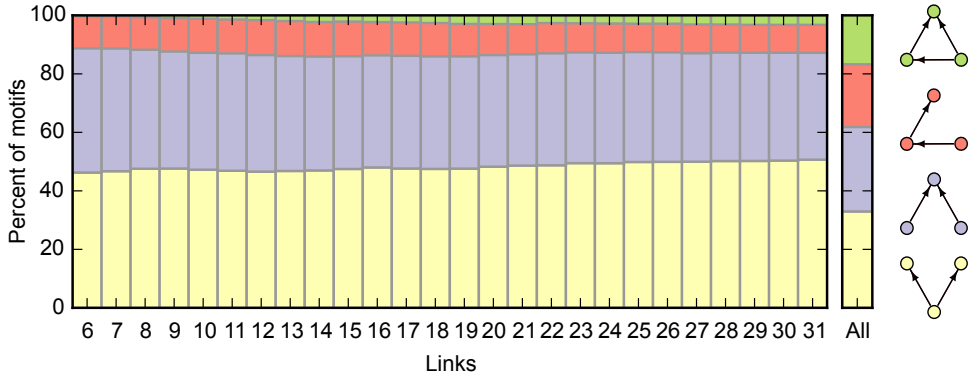


Figure 24: Motifs analysis of the backbones of interactions. Each bar of the left plot describes the motif representation of the different k-link backbones of interactions found for every network in our dataset. The bar in the right plot describes the motif representation of the entire networks. The different colors characterize the proportion of each of the motifs found, which are represented on the right side of the figure. Notice that for the study of the motif structure of the entire networks, other motifs were found but in very low proportions, which make them imperceptible.

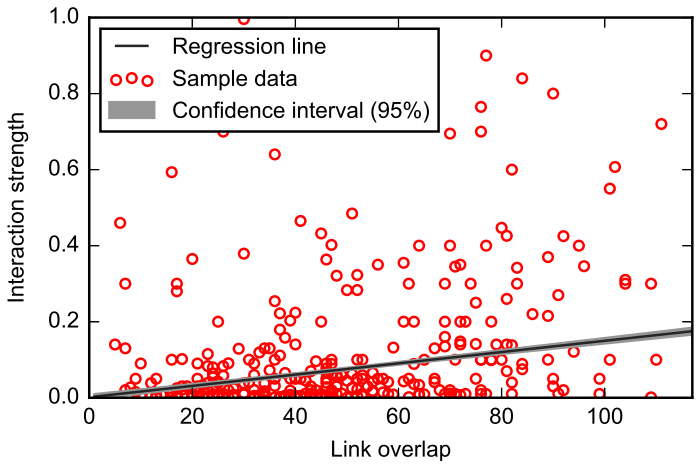


Figure 25: Relationship between interaction strength and link overlap. The red dots are a random sample of the all the data. The black line characterizes the result a linear regression of the data ($F_{1,16610} = 818.1$, $p < 0.01$) and the gray area shows the corresponding 95% confidence intervals.

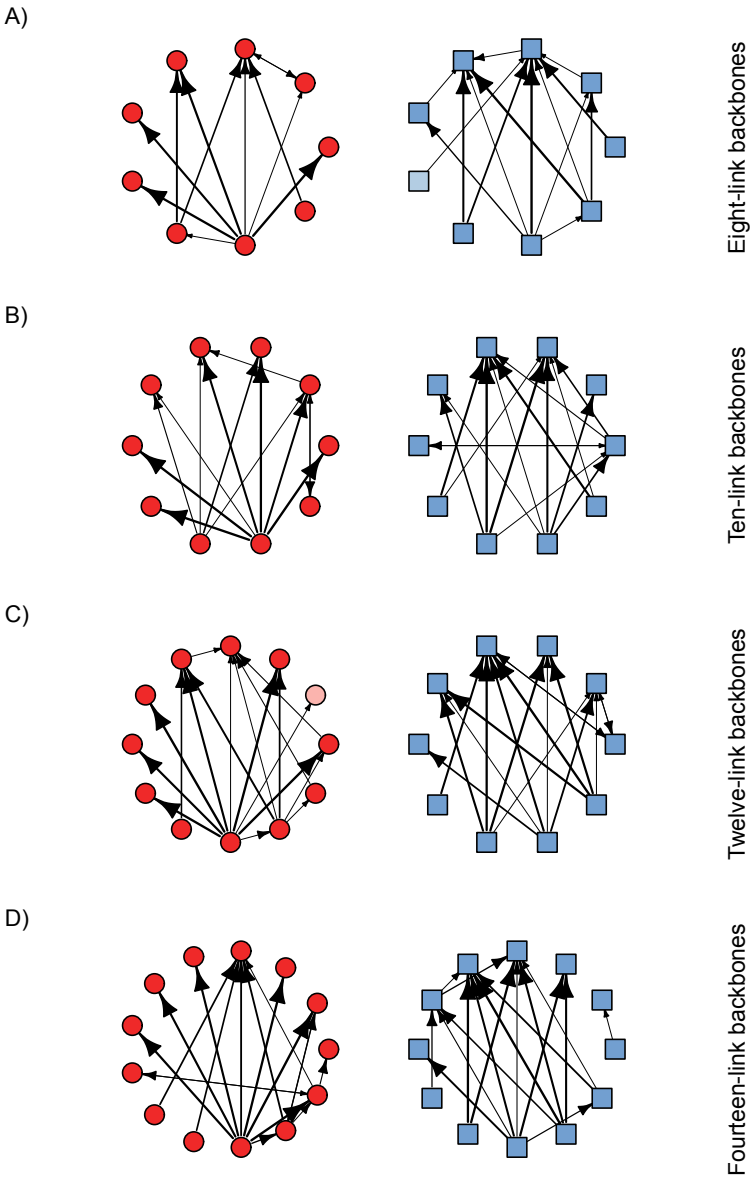


Figure 26: Visualization of different-sized k-link backbones of interactions found across all food webs: (A) structures characterizing the 8-link backbones of interactions; (B) structures characterizing the 10-link backbones of interactions; (C) structures characterizing the 12-link backbones of interactions; and (D) structures characterizing the 14-link backbones of interactions. These structures are found by selecting the medoids of the clusters and overlapping them with all the within-cluster backbones, following the example shown in Fig. 1C of the main text (see also Fig. 4 from the main text). The weight of the links is proportional to the likelihood l of finding them in the backbones, and the light-shaded nodes represent nodes that significantly appear in the k-link backbones but not in the medoids. Note that links that were not significantly represented in the backbones ($l < 0.01$) are not shown.

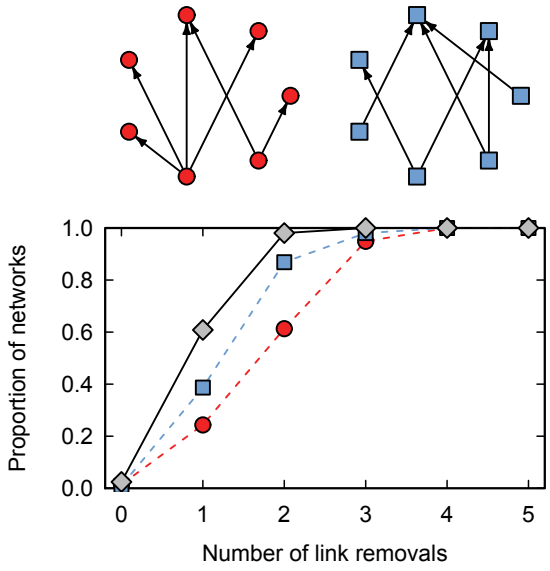


Figure 27: Visualization of the backbones of interaction shared across food webs. The two networks depicted in this plot are the structures that encompass the greatest variation across our food-web dataset and can explain over 60% of the backbones within one link removal (see Supplementary Note 1). The dashed lines show the proportion of networks explained by each of these structure individually whereas the gray diamonds show the proportion of networks that those structure can explain in combination.

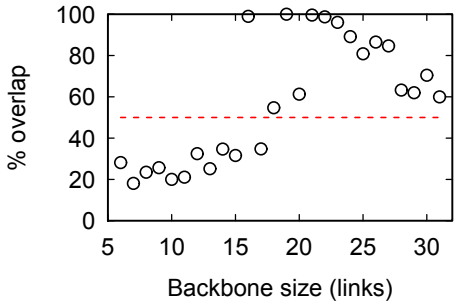


Figure 28: Analysis of the quality of the backbone clustering. Every point represents the overlap of the 95% confidence ellipses that characterize the two clusters found for any given backbone size k. The red dotted line represents the 50% ellipse overlap.

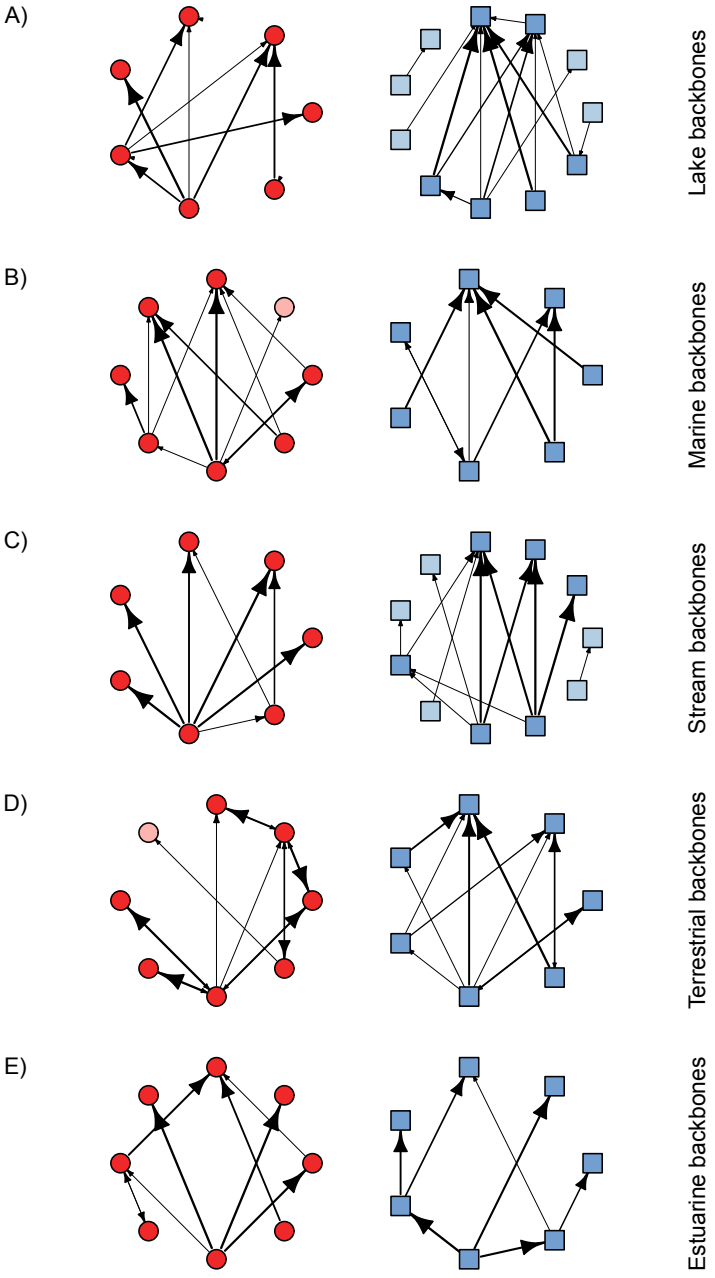


Figure 29: Visualization of the 6-link backbones of interactions found across food webs from the different ecosystem types: (A) lake backbones; (B) marine backbones; (C) stream backbones; (D) terrestrial backbones; and (E) estuarine backbones. These structures are found by selecting the medoids of the clusters and overlapping them with all the within-cluster backbones, following the example shown in Fig. 1C of the main text (see also Fig. 4 from the main text). The weight of the links is proportional to the likelihood l of finding them in the backbones, and the light-shaded nodes represent nodes that significantly appear in the k-link backbones but not in the medoids. Note that links that were not significantly represented in the backbones ($l < 0.01$) are not shown.

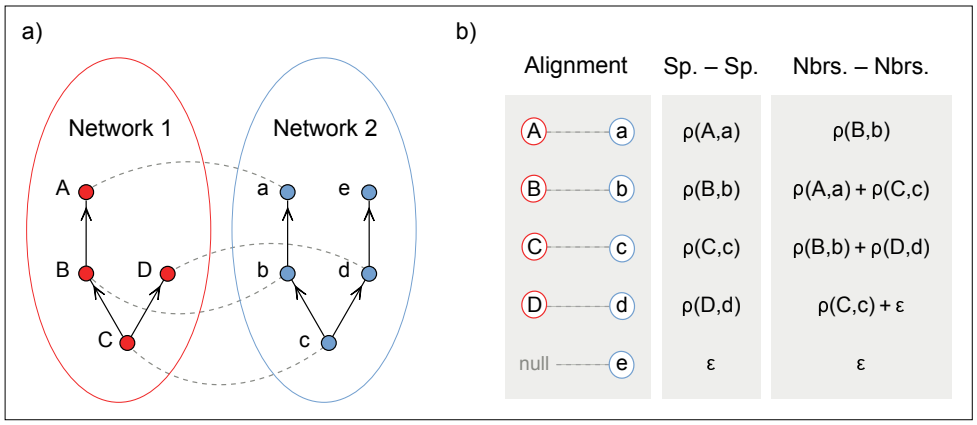


Figure 30: Example of the pairwise role distance computation when pairing species to species and species' neighbors to species' neighbors. (a) Example of an alignment between two simple networks. Notice that in this example we consider one of the best possible alignments between the example networks. (b) A description of the pairwise distance between paired species according to both a species-species pairing and a neighbors-neighbors pairing. In this figure, $\rho(i, j)$ represents the role distance between species i and j, and ϵ characterizes the default contribution of a unpaired species.

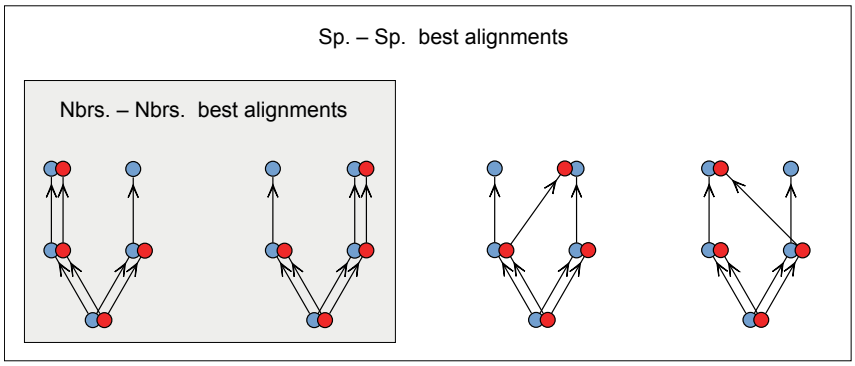


Figure 31: Best possible alignments between networks 1 and 2 from Supplementary Fig. 30. Pairing species to species can be done by optimizing Eq. 5 of the main text, resulting into 4 indistinguishable alignments. Pairing species' neighbors to species' neighbors requires optimizing Eq. 5 of the main text and only produce 2 indistinguishable alignments.

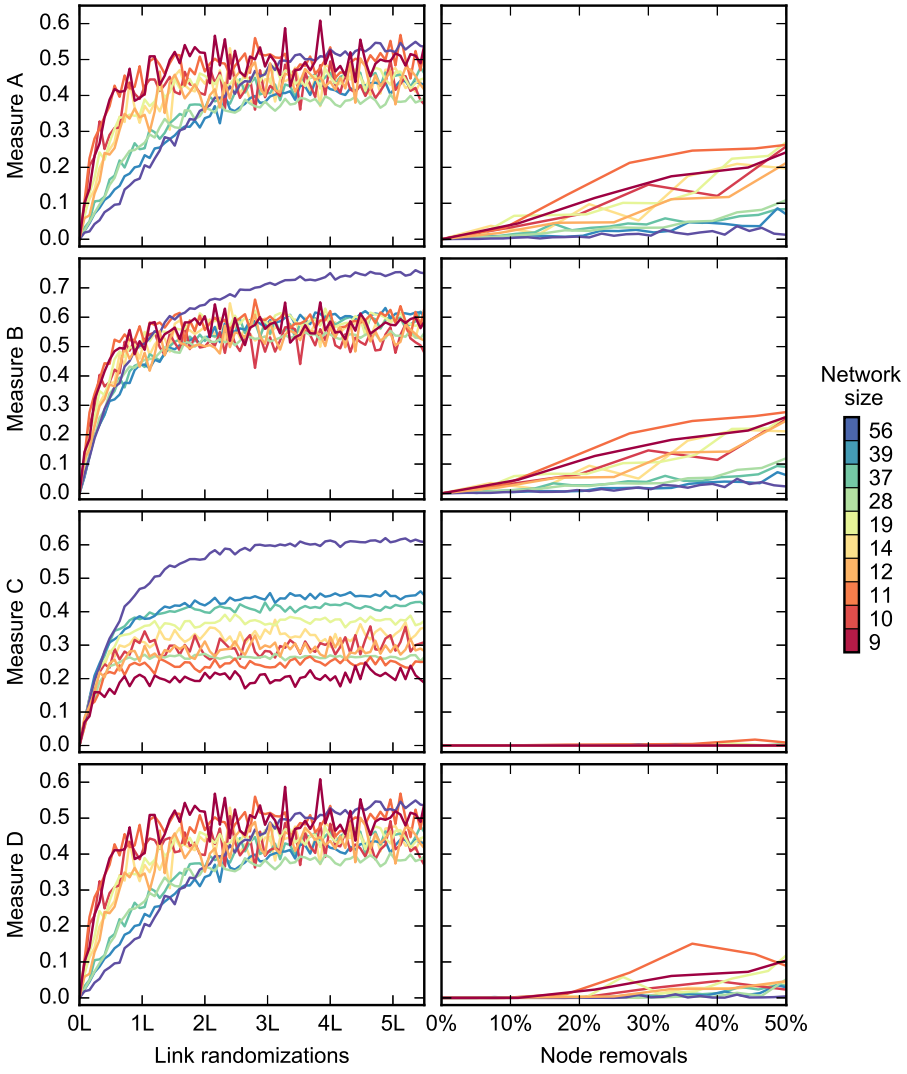


Figure 32: Tests for the alignment algorithm and alignment quality measures presented in the previous section. The left panels show the results obtained when aligning the networks to random versions of themselves to test the effect of link differences between aligned networks. The number of link randomizations is expressed as a multiple of the total number of links L of each network. The right panels show the results obtained when comparing networks to reduced versions of themselves to test the effect of size difference between aligned networks. The colors of the different lines are chosen based on the size of the networks.

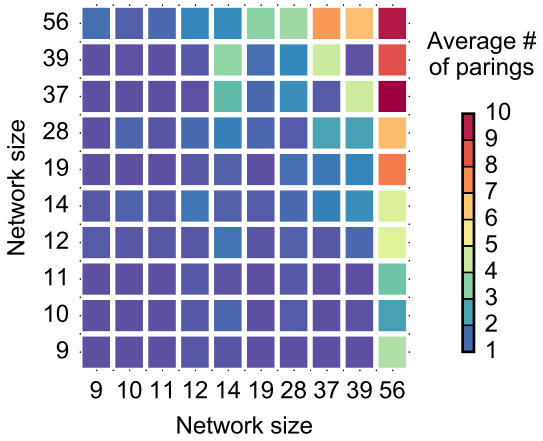


Figure 33: Analysis of the alignment variability across networks. The panel shows the results obtained when optimally aligning 100 times every pair of networks used in the previous section. The color characterizes the average number of parings of any species of a given network. The networks are sorted based on their size (i.e. number of species).

		Measures of alignment quality			
		A	B	C	D
PERMANOVA	Overall	22.81	24.27	27.20	24.99
	Size-constrained	5.01	3.49	-1.06	4.87

Table 2: F-test statistic obtained using the PERMANOVA test when comparing the alignments between networks from 5 different ecosystems. In the overall analysis we compare 411 networks whereas in the size-constrained analysis we compare only 116 networks. The p-values found for each of these comparisons were $p < 0.01$ with the exception of the size-constrained statistical test performed on the size-constrained dissimilarity matrix (red value).

	Estuarine	Lake	Marine	Stream	Terrestrial
Estuarine		4.24	18.43	14.93	0.76
Lake			6.19	36.42	8.01
Marine				91.22	18.19
Stream					18.30

Table 3: F-test statistic obtained using the PERMANOVA test when comparing network alignments between networks from different pairs of ecosystem types. In this case, we used the alignment quality measure A to estimate the dissimilarity matrix. The p-values found for each of these comparisons were $p < 0.01$ with the exception of the comparison between estuarine and terrestrial food webs that was not statistically significant (red value).

BIBLIOGRAPHY

Marti J Anderson. A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26(1):32–46, 2001.

Elizabeth A Leicht and Mark EJ Newman. Community structure in directed networks. *Physical Review Letters*, 100(11):118703, 2008.

Daniel B Stouffer and Jordi Bascompte. Compartmentalization increases food-web persistence. *Proceedings of the National Academy of Sciences*, 108(9):3648–3652, 2011.

Daniel B Stouffer, Marta Sales-Pardo, M Irmak Sirer, and Jordi Bascompte. Evolutionary conservation of species' roles in food webs. *Science*, 335(6075):1489–1492, 2012.

D

SUPPLEMENTARY INFORMATION S4

Supporting information for Chapter 4: Untangling the seasonal dynamics of plant-pollinator communities.

D.1 Supplementary Methods

Species' structural roles

There are multiple ways to characterize the structural roles of species in a community. In this work, we used a definition based on the idea of network motifs Milo et al. (2002), which characterize the distinct patterns of interactions between n species found within a network. Following the ideas presented by Stouffer et al. (2012) and (Baker et al., 2015), we defined the role of any species α based on the number of times it appears in any of the unique positions of the n-species network motifs. In particular, we used the tools developed by (Bramon Mora et al., 2018) to describe the role of α as measure of the number of times α appears in any of the distinct positions found within motifs made of 3, 4 and 5 species. Importantly, this measure is designed in such a way that incorporates information regarding the interaction strengths between species (Bramon Mora et al., 2018).

Alignment matrix

We compiled all alignments between the networks in our dataset using the alignment matrix M. In particular, M contains information regarding every pairwise alignment between the networks $\{A, B, C, D...\}$ in our dataset—in our case, this corresponds to the pairwise alignment between 44 empirical networks. Given a set of n alignments $\{\lambda\}$ between A and B, we calculate a given element m_{ij}^{AB} as $\sum_{\lambda}^{n} k_{\lambda}$, where k_{λ} is 0 if species $i \in A$ and $j \in B$ are not paired in the alignment λ, and c_{ij} otherwise. Notice that c_{ij} characterizes the quality of the pairing i-j between $i \in A$ and $j \in B$ as described in the main text.

Species dynamics across groups: constant transition probabilities

Given k groups of positions, we considered a system in which any species can be found in $k+2$ possible states $\vec{y}(t)$ at time t. These include: a state y_s describing a species that has not yet entered the

network; k states y_k characterizing a species in each of the different groups of positions; and, a state y_e describing a species that has already exited the network. Then, we used a Bayesian multinomial logistic regression as a way to estimate $\Pr(\vec{y}(t)|\vec{y}(t-\Delta t))$, the transition probabilities across the different states of the system over time. In particular, we used $k+1$ linear models to estimate these constant transition probabilities (McElreath, 2018). Importantly, we assumed $\Pr(y_k|y_e) = 0$ and $\Pr(y_s|y_e) = 0$. In addition, we also assumed species phenologies to be uninterrupted; therefore, we considered any observation of a species transition from any state y_k to y_s to be a detection error. Following this, we ignored any observation of species moving from any y_k to y_s, and the subsequent transition from y_s to any y_k.

Species dynamics across groups: density-dependent transition probabilities

In this case, we used a Bayesian multinomial logistic regression as a way to estimate the transition probabilities $\Pr(\vec{y}(t)|\vec{y}(t-\Delta t), N_{t-\Delta t})$, where $N_{t-\Delta t}$ describe the total number of species at time $N_{t-\Delta t}$. In particular, we used $k+1$ linear models to estimate these density-dependent transition probabilities (McElreath, 2018).

D.2 Supplementary Figures and Tables

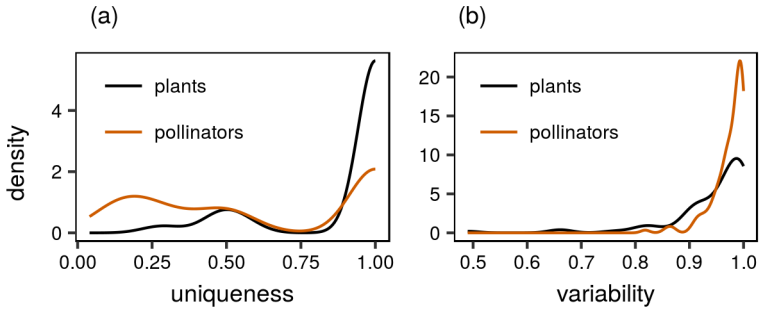


Figure 34: Analysis of species' positions across binary networks. (a) Uniqueness of the position of plant and pollinator species within networks. (b) Variability of species' position across networks, including only plant and pollinator species that appear in multiple networks during a season.

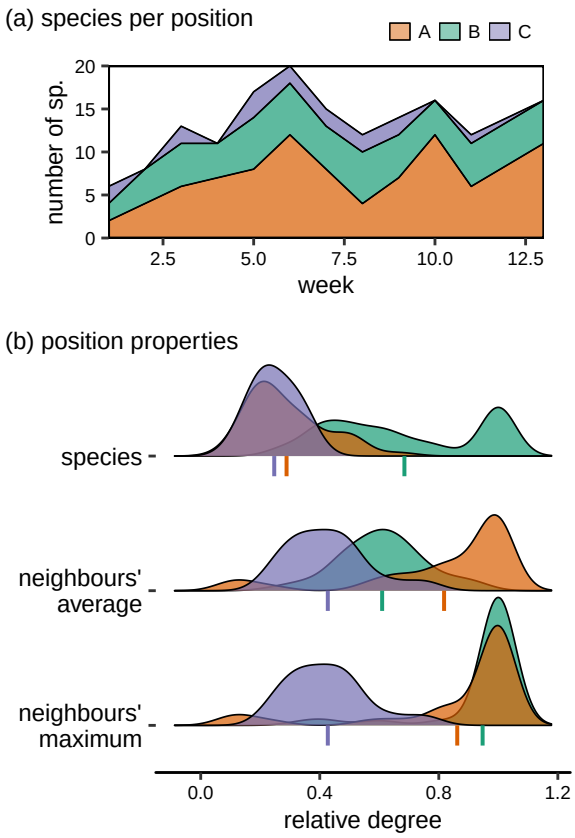


Figure 35: Distinguishing properties of the different groups of pollinator positions found for the 2013 sampling season. (a) Number of pollinator species in each group of positions over time. Each color represents a different group of positions. (b) Different properties summarizing the species forming each group. The first panel depicts the relative degree distribution of the best connected neighbor of every species in each group. The second panel shows the neighbours' average relative degree for the species in each group. The third panel shows the relative degree distribution of species in each group. The colored segments depicted under the distributions characterize the mean of each distribution.

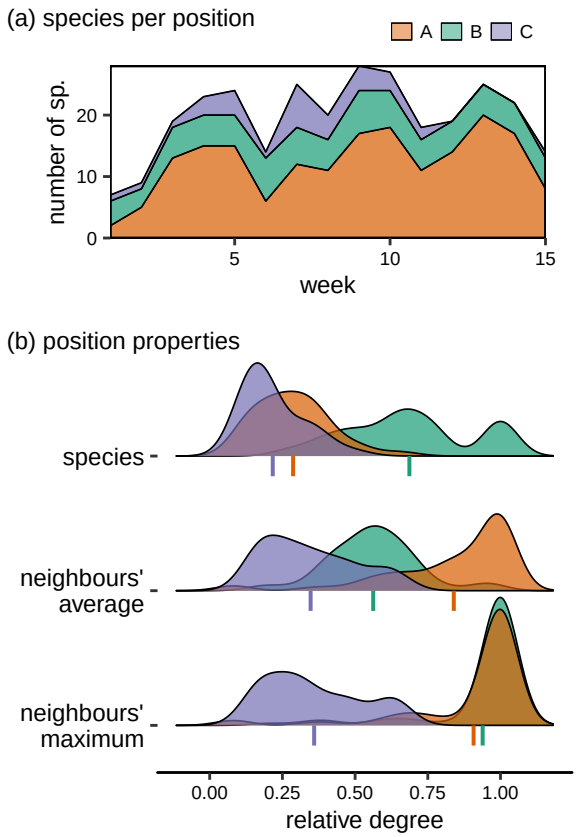


Figure 36: Distinguishing properties of the different groups of pollinator positions found for the 2014 sampling season. (a) Number of pollinator species in each group of positions over time. Each color represents a different group of positions. (b) Different properties summarizing the species forming each group. The first panel depicts the relative degree distribution of the best connected neighbor of every species in each group. The second panel shows the neighbours' average relative degree for the species in each group. The third panel shows the relative degree distribution of species in each group. The colored segments depicted under the distributions characterize the mean of each distribution.

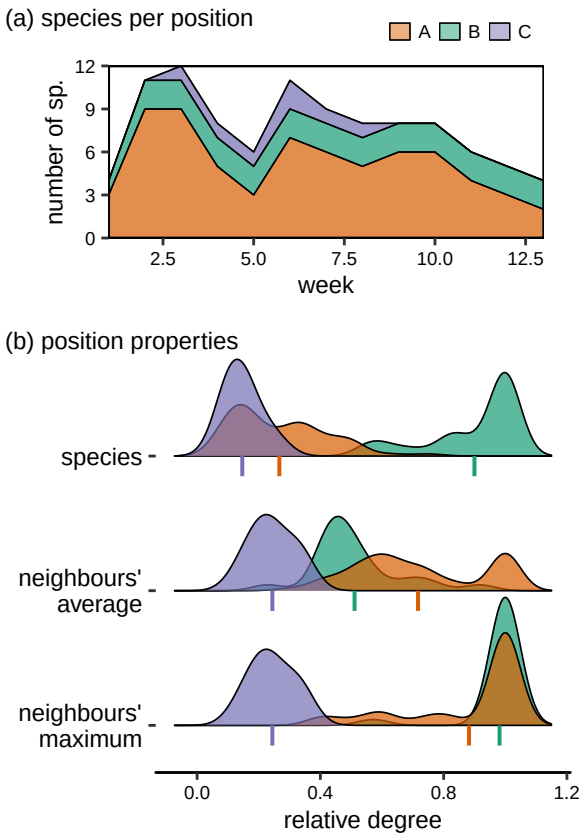


Figure 37: Distinguishing properties of the different groups of plant positions found for the 2013 sampling season. (a) Number of plant species in each group of positions over time. Each color represents a different group of positions. (b) Different properties summarizing the species forming each group. The first panel depicts the relative degree distribution of the best connected neighbor of every species in each group. The second panel shows the neighbours' average relative degree for the species in each group. The third panel shows the relative degree distribution of species in each group. The colored segments depicted under the distributions characterize the mean of each distribution.

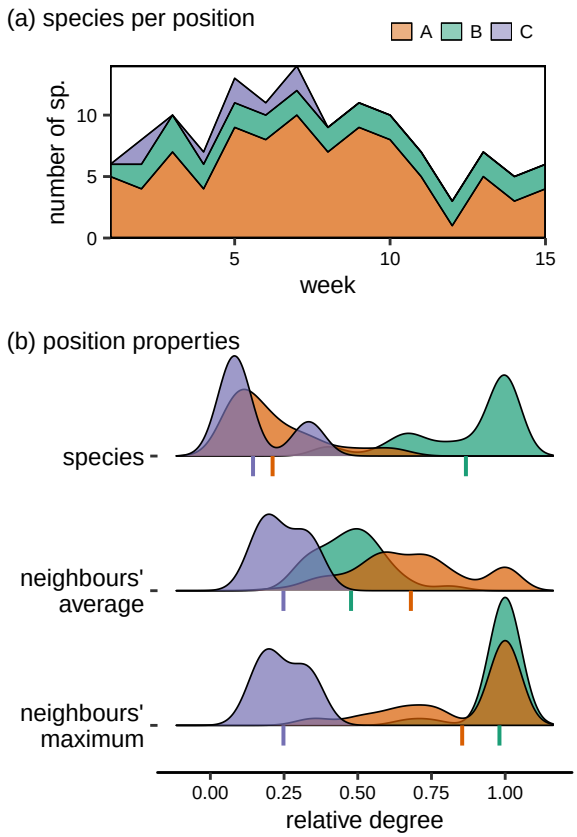


Figure 38: Distinguishing properties of the different groups of plant positions found for the 2014 sampling season. (a) Number of plant species in each group of positions over time. Each color represents a different group of positions. (b) Different properties summarizing the species forming each group. The first panel depicts the relative degree distribution of the best connected neighbor of every species in each group. The second panel shows the neighbours' average relative degree for the species in each group. The third panel shows the relative degree distribution of species in each group. The colored segments depicted under the distributions characterize the mean of each distribution.

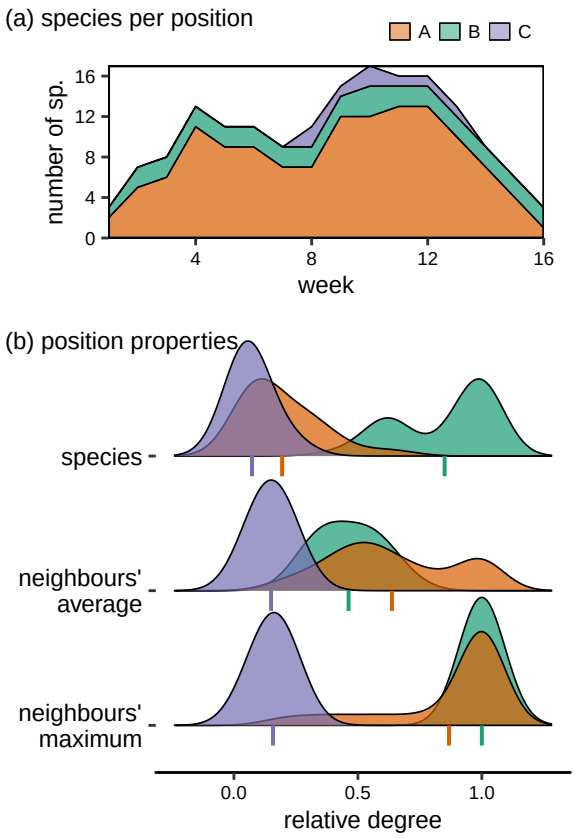


Figure 39: Distinguishing properties of the different groups of plant positions found for the 2014 sampling season. (a) Number of plant species in each group of positions over time. Each color represents a different group of positions. (b) Different properties summarizing the species forming each group. The first panel depicts the relative degree distribution of the best connected neighbor of every species in each group. The second panel shows the neighbours' average relative degree for the species in each group. The third panel shows the relative degree distribution of species in each group. The colored segments depicted under the distributions characterize the mean of each distribution.

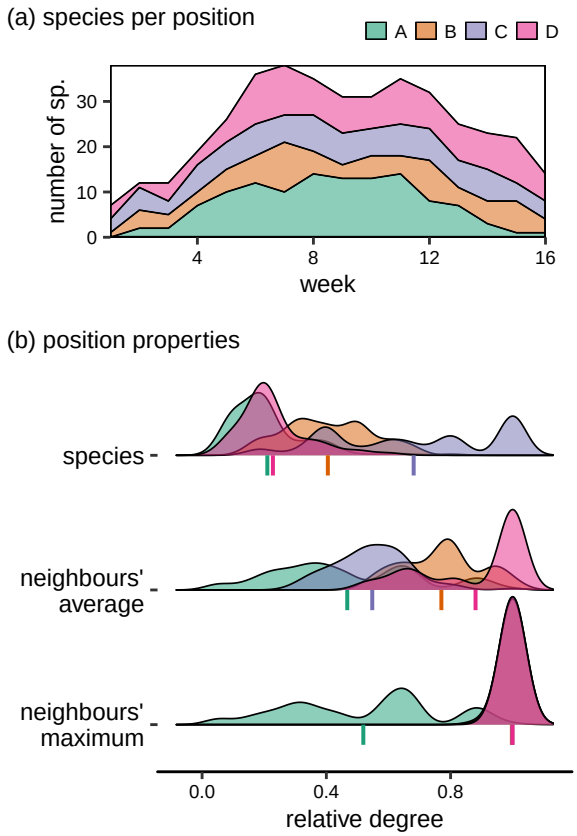


Figure 40: Distinguishing properties of the different groups of species' position found for the 2015 sampling season using the multi-level community detection method (Table 1; Blondel et al. 2008). (a) Number of pollinator species in each group of positions over time. Each color represents a different group of positions. (b) Different properties summarizing the species forming each group. The first panel depicts the relative degree distribution of the best connected neighbor of every species in each group. The second panel shows the neighbours' average relative degree for the species in each group. The third panel shows the relative degree distribution of species in each group. The colored segments depicted under the distributions characterize the mean of each distribution.

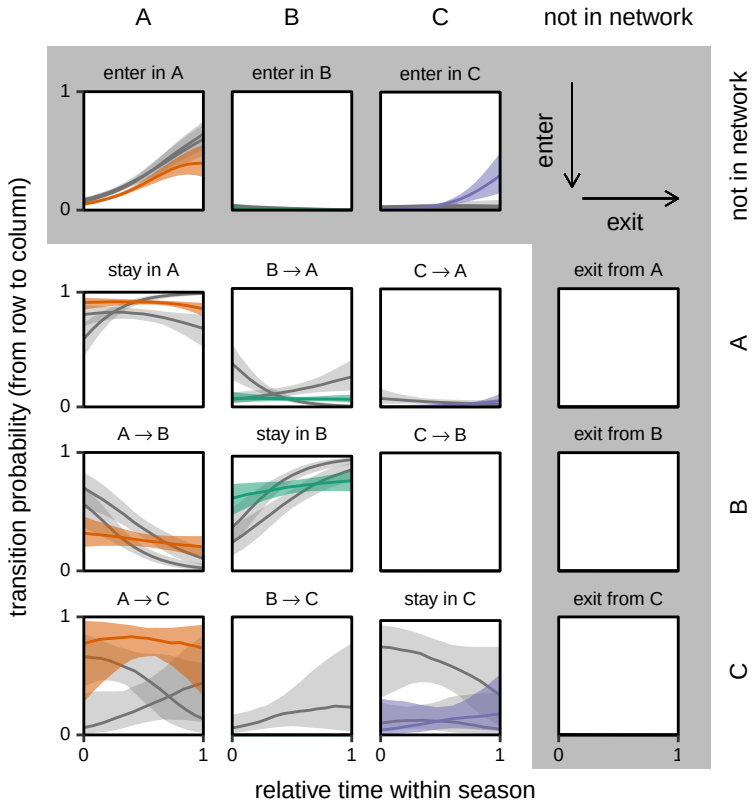


Figure 41: Movement of plant species across position groups. Every panel in the matrix describes the transition probability between the different groups of positions. The order of the matrix is such that it characterizes the transition probabilities from row groups to column groups over time. The different groups are those presented in Supplementary Figs 37–39. The shaded row describes the probabilities of species entering the network into the different groups (A, B and C) whereas the shaded column describes the probabilities of species exiting the network from each of these groups. The different lines in the graph represent the results found for every sampling year, where the shade of each line characterizes the first and third quantiles. The colored lines indicate the results estimated for the 2015 sampling season. Notice that the color is chosen based on the recipient group of the transition probability.

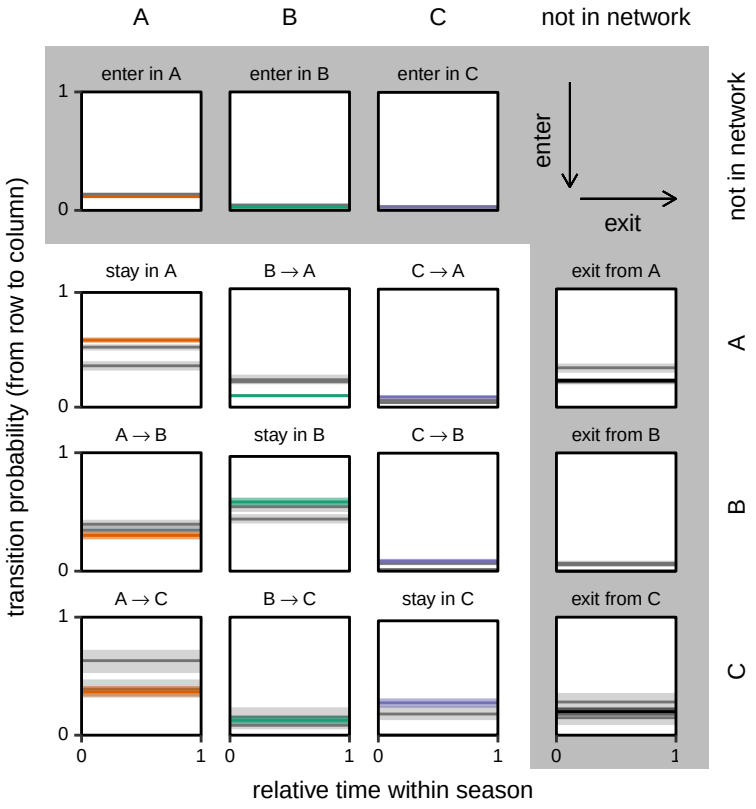


Figure 42: Movement of pollinator species across position groups estimated using a model with constant transition probabilities (Supplementary Methods). Every panel in the matrix describes the transition probability between the different groups of positions. The order of the matrix is such that it characterizes the transition probabilities from row groups to column groups over time. The different groups are those presented in Fig. 3 of the main text. The shaded row describes the probabilities of species entering the network into the different groups (A, B and C) whereas the shaded column describes the probabilities of species exiting the network from each of these groups. The different lines in the graph represent the results found for every sampling year, where the shade of each line characterizes the first and third quantiles. The colored lines indicate the results estimated for the 2015 sampling season. Notice that the color is chosen based on the recipient group of the transition probability.

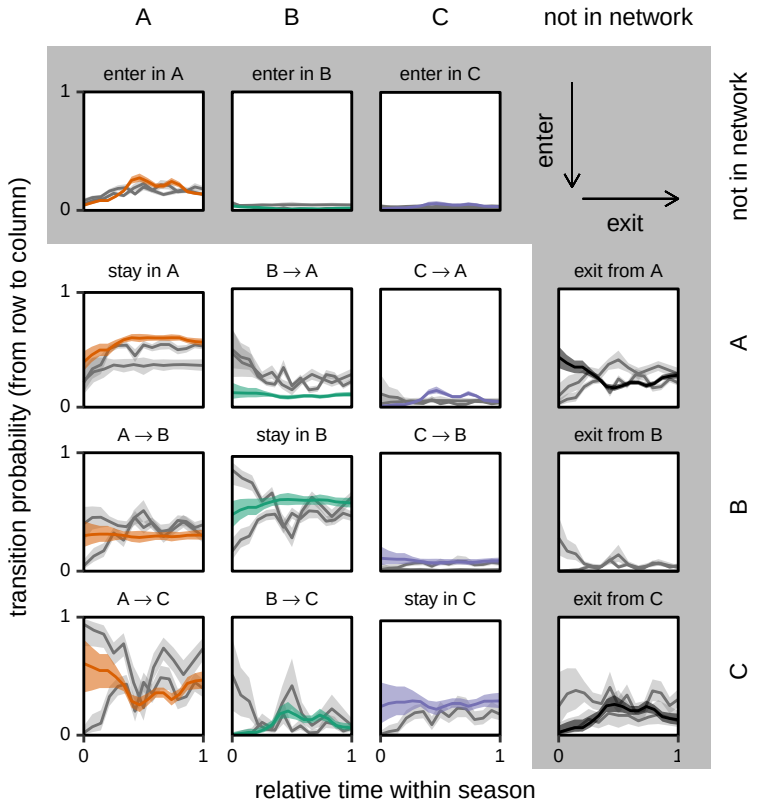


Figure 43: Movement of pollinator species across position groups estimated using a model with density-dependent transition probabilities (Supplementary Methods). Every panel in the matrix describes the transition probability between the different groups of positions. The order of the matrix is such that it characterizes the transition probabilities from row groups to column groups over time. The different groups are those presented in Fig. 3 of the main text. The shaded row describes the probabilities of species entering the network into the different groups (A, B and C) whereas the shaded column describes the probabilities of species exiting the network from each of these groups. The different lines in the graph represent the results found for every sampling year, where the shade of each line characterizes the first and third quantiles. The colored lines indicate the results estimated for the 2015 sampling season. Notice that the color is chosen based on the recipient group of the transition probability.

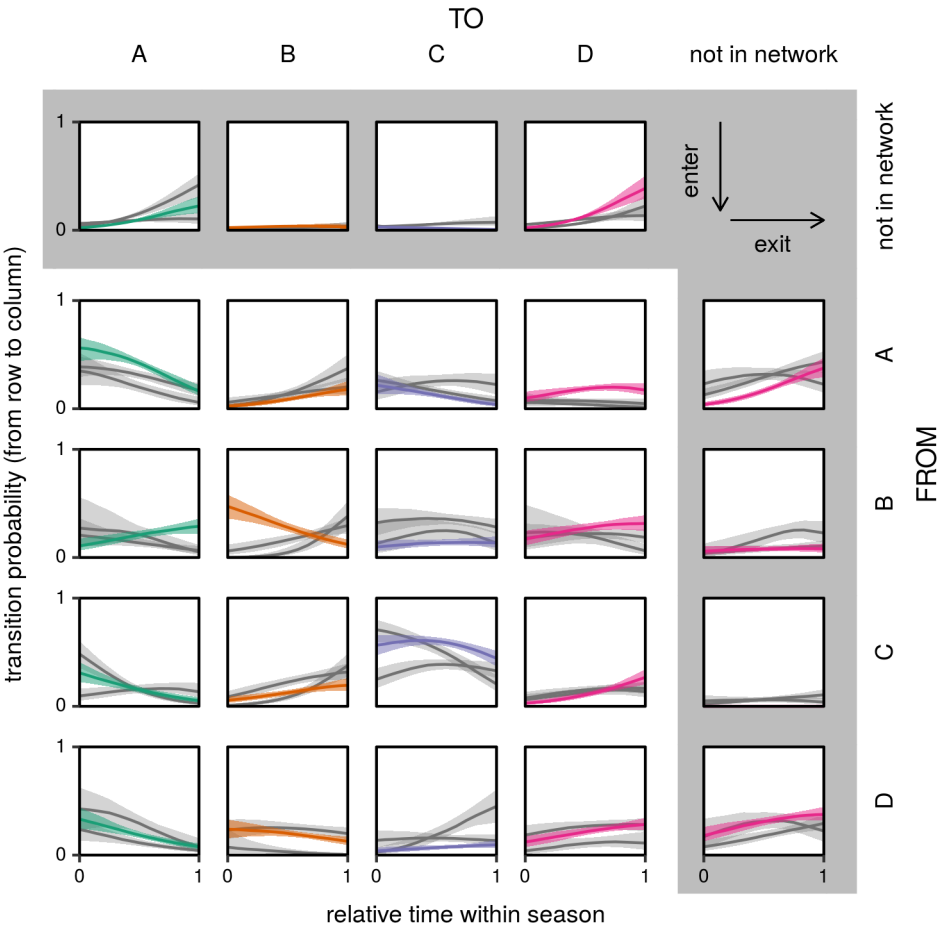


Figure 44: Movement of pollinator species across position groups. Every panel in the matrix describes the transition probability between the different groups of positions. The order of the matrix is such that it characterizes the transition probabilities from row groups to column groups over time. The different groups are those presented in Supplementary Fig. 40. The shaded row describes the probabilities of species entering the network into the different groups (A, B, C and D) whereas the shaded column describes the probabilities of species exiting the network from each of these groups. The different lines in the graph represent the results found for every sampling year, where the shade of each line characterizes the first and third quantiles. The colored lines indicate the results estimated for the 2015 sampling season. Notice that the color is chosen based on the recipient group of the transition probability.

			Normalized mutual information				
	pol.	pl.	*greedy*	*betweenness*	*random walk*	*infomap*	*louvain*
greedy	2	2	1.00	0.69	0.66	0.59	0.67
betweenness	2	2	0.69	1.00	0.68	0.60	0.62
random walk	3	3	0.66	0.68	1.00	0.74	0.74
infomap	7	3	0.59	0.60	0.74	1.00	0.78
louvain	4	2	0.67	0.62	0.73	0.78	1.00

Table 4: Community detection methods summary. The first column lists the community detection methods used here. The methods used are the following: greedy optimization (greedy; Clauset et al. 2004); edge betweenness (betweenness; Newman and Girvan 2004); short random walks Pons and Latapy 2006; map equation optimization (infomap; Rosvall et al. 2009); and, multi-level optimization (louvain; Blondel et al. 2008). The second and third columns describe the number of pollinator (pol.) and plant (pl.) groups found with each method. The rest of the columns provide a pairwise comparison of the results found with the different methods by means of the normalized mutual information criteria (Danon et al., 2005). The methods highlighted are the ones used in the paper.

BIBLIOGRAPHY

Nick J Baker, Riikka Kaartinen, Tomas Roslin, and Daniel B Stouffer. Species' roles in food webs show fidelity across a highly variable oak forest. *Ecography*, 38(2):130–139, 2015.

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008 (10):P10008, 2008.

Bernat Bramon Mora, Alyssa R Cirtwill, and Daniel B Stouffer. pymfinder: a tool for the motif analysis of binary and quantitative complex networks. *bioRxiv*, page 364703, 2018.

Aaron Clauset, Mark EJ Newman, and Cristopher Moore. Finding community structure in very large networks. *Physical Review E*, 70 (6):066111, 2004.

Leon Danon, Albert Diaz-Guilera, Jordi Duch, and Alex Arenas. Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(09):P09008, 2005.

Richard McElreath. *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC, 2018. pp. 323–326.

Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.

Mark EJ Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2):026113, 2004.

Pascal Pons and Matthieu Latapy. Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications*, 10(2):191–218, 2006.

Martin Rosvall, Daniel Axelsson, and Carl T Bergstrom. The map equation. *The European Physical Journal Special Topics*, 178(1):13–23, 2009.

Daniel B Stouffer, Marta Sales-Pardo, M Irmak Sirer, and Jordi Bascompte. Evolutionary conservation of species' roles in food webs. *Science*, 335(6075):1489–1492, 2012.